Assembly and analysis of a comprehensive phosphotyrosine-dependent protein-protein interaction network

Dissertation
zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)
im Fach Biologie
eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von Herrn Dipl.-Biol. Arndt Großmann

Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Jan-Hendrik Olbertz
Dekan der Lebenswissenschaftlichen Fakultät
Prof. Dr. Richard Lucius

Gutachter: Prof. Dr. Christian Spahn
Prof. Dr. Erich Wanker
Prof. Dr. Hans Lehrach

Tag der mündlichen Prüfung: 27.11.2015

Acknowledgements

First and foremost, I want to thank Prof. Dr. Uli Stelzl for allowing me to carry out this project under his supervision and providing the environment necessary for such a large-scale endeavor. I truly appreciate his continued support throughout the project. I believe that receiving regular feedback from a competent and open-minded discussion partner was not only important for the quality of this study, but also for my personal development.

I am grateful to Nouhad Benlasfer, who established our mammalian cell culture and carried out most of the related experiments presented in this study. What is more, she answered all my questions about the technical aspects of working with mammalian cells, with hands-on instructions where appropriate. This allowed me to learn a lot very quickly.

I am thankful to Thomas Przewieslik, the engineer who almost single-handedly assembled our stamping robot, and Petra Birth, who conducted a substantial amount of yeast two-hybrid experiments for this project.

I am also grateful to Prof. Dr. Christian Spahn for agreeing to be my supervisor for the purposes of getting a degree from the Humbold University, to Prof. Dr. Edda Klipp for chairing my dissertation committee, to Prof Dr. Erich Wanker and Prof. Dr. Hans Lehrach for writing reports and to Prof. Dr. Peter Hammerstein, Prof. Dr. Hanspeter Herzel and Prof. Dr. Nils Blüthgen for being members of the dissertation committee. In addition, Prof Dr. Erich Wanker and his group were excellent collaboration partners, in the context of the NeureNet project and beyond.

Finally, I want to thank my friends and family for supporting me during this time.

Abstract

Protein-protein interactions govern cellular functions on the molecular level. Post-translational modifications alter these interactions allowing highly dynamic regulation. Protein tyrosine phosphorylation is an especially relevant post-translational modification, because it is tightly linked to intercellular regulation of growth and development in metazoans. Diseases like cancer or autoimmune disorders arise from misregulation of these processes generating great medical interest in protein tyrosine phosphorylation and processes relating to it. This study provides a comprehensive set of 292 mostly novel, high-quality phosphotyrosine-dependent protein-protein interactions detected in genome-scale yeast two-hybrid screens using full-length proteins filling a gap in phosphotyrosine signaling knowledge, which has so far been based largely on peptide binding and affinity purification-coupled mass spectrometry experiments. The high quality was demonstrated experimentally and computationally, in co-immunoprecipitation and protein complementation assays, as well as over-representation analyses and comparison to prior knowledge. Previously reported linear peptide motifs are reflected in the binding partners, but clearly do not account for most of the interactions, emphasizing the relevance of full-length protein context. The interactions were further shown to form an unusually dense, monolithic network with a central core and reflect and expand phosphotyrosine-related KEGG pathways. Seven of the eight core proteins are well-established signaling hubs. The eighth core gene, SH2D2A, seems to play a more central role than currently appreciated. Finally, selected interactions involving GRB2 were shown to occur in different specific subcellular localizations. Together, these results strongly suggest that the interactions presented here represent an important step toward understanding growth and development and will benefit treatment of pressing medical issues substantially.

Zusammenfassung

Protein-Protein-Wechselwirkungen steuern zelluläre Funktionen auf molekularer Ebene. Posttranslationale Proteinmodifikationen beeinflussen diese Wechselwirkungen und erlauben dynamische Regulierung. Tyrosinphosphorylierung ist eine besonders relevante Modifikation, weil sie eng mit interzellulärer Regulation von Wachstum und Enticklung in Vielzellern verbunden ist. Da falsche Regulierung dieser Prozesse zu Krebs oder Autoimmunerkrankungen führen kann, ist sie auch von großem medizinischen Interesse. In Hefe-Zwei-Hybrid-Untersuchungen mit Volllängen-Proteinen im Genommaßstab wurde ein umfassender Satz von 292 größtenteils neuen phosphotyrosinabhängigen Proteinwechselwirkungen erster Güte ermittelt. Damit wurde eine Wissenslücke im Bereich der phosphotyrosinabhängigen Signalübertragung, der bisher hauptsächlich auf Peptidbindungs- und Affinitätsaufreinigungs-gekoppelten Massenspektronomieexperimenten fußte. Die Güte der Interaktionen wurde experimentell und informatisch, in Coimmunpräzipitations- und Proteinkomplementierungs-, sowie in Überrepräsentationsanalysen und Literaturvergleichen, gezeigt. Bekannte lineare Bindesequenzmotive kommen zwar gehäuft vor, können die Mehrzahl der Interaktionen aber offensichtlich nicht erklären. Die Wechselwirkungen bilden ein dichtes, einheitliches Netzwerk und widerspiegeln phosphotyrosinabhängige KEGG-Signalwege. Es hat ein Herzstück aus acht Genen, von denen sieben fest etablierte Signalverarbeitungshauptknotenpunkte darstellen. Dem achten, SH2D2A, scheint eine deutlich wichtigere Rolle zuzukommen als bisher wahrgenommen. Schliesslich wurde für eine Auswahl von GRB2-Interaktionen unterschiedliche subzelluläre Verortung vorgenommen. Zusammengenommen legen diese Ergebnisse nah, dass die hier veröffentlichten Wechselwirkungen einen wesentlichen Schritt für das Verstehen von Wachstum und Entwicklung markieren und zur Verbesserung der Behandlungsmöglichkeiten in wichtigen Medizinbereichen beitragen werden.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

More than a decade after the sequencing of the human genome (Venter et al. 2001), many of its promises are still not realized (Wade 2010). The main reason is that while the genome sequence may, in principle, contain all the information relevant for the creation and operation of an organism, most of the cells functions are realized by proteins (Cusick et al. 2005). This is especially true for evolutionarily new functions (Joyce 2002). These functions are often more relevant in a medical context, because the older functions, like splicing or translation, are so fundamental that defects are often lethal. In fact, human diseases usually do not arise from defects in catalytic proteins but in regulatory proteins (Cuccato et al. 2009). Most regulatory protein functions emerge from the interaction of several proteins (Charbonnier et al. 2008). Therefore, understanding human cells requires the knowledge of protein-protein interactions (Stelzl & Wanker 2006). In the recent past, several large-scale protein-protein interaction studies have started working towards a complete map of all human protein-protein interactions, the so-called interactome (Stelzl et al. 2005; Rual et al. 2005).

Complex organisms require complex control mechanisms. This is reflected in their composition. Usually, increases in compartmentalization, genome size, number of different cell types, organism size, phenotypic diversity, and epigenetic mechanisms like alternative splicing and post-translational modification of proteins are correlated (Schad et al. 2011). Post-translational modifications, in particular, play an important role, because they allow efficiently switching protein interactions on or off (Yang 2005; Deribe et al. 2010).

1.1 Post-translational protein modifications affect protein binding

Physical interactions between proteins are fundamental for the control of cellular processes. It is important to note that protein-protein interactions are not intrinsically binary, but (simplifyingly) thinking about them in that way is extremely fruitful. Protein-protein interactions form a network that contains all the possibilities for information transfer, especially regulation, but is incomplete in that it is a map only (Stelzl & Wanker 2006). In practice, protein-protein interactions depend not only on the proteins fundamental capacity for interaction, but also on protein expression, localization and state. The most efficient way of altering the protein-protein interaction network states is by altering protein states through covalent modification. In contrast to other possibilities, like raising or decreasing expression or degradation rates, post-translational protein modifications can very efficiently be coupled to energetically favorable reactions, like hydrolysis of ATP, and therefore achieve very high turnover rates.

1.2 Phosphotyrosine is a post-translational protein modification strongly linked to multicellularity

Phosphotyrosine signaling is intimately linked to the regulation of cellular processes related to multicellularity and overcoming the challenges it entails, as well as the diseases that arise when these processes are deregulated (Suga et al. 2014).

1.2.1 Regulation of growth and development is related to phosphotyrosine signaling

For individual cells to form successful multicellular organisms, they have to find a way of agreeing on a course of action (or inaction) and abide by it (Kaiser 2001). Above all, growth and development have to be coordinated closely. In

practice, much of this kind of communication is brought about by proteins spanning the plasma membrane and, upon binding an extracellular peptide ligand, effecting phosphorylation of tyrosine residues on intracellular peptide chains, in cooperation with downstream proteins. The extracellular ligand can be transmembrane proteins belonging to neighboring cells, but most are soluble. These peptides are called growth factors and their receptors are called receptor tyrosine kinases (Ullrich & Schlessinger 1990).

1.2.1.1 Growth factors activating receptor tyrosine kinases are exemplary for phosphotyrosine signaling

The binding of growth factors to receptor tyrosine kinases and subsequent activation of receptor tyrosine kinase pathways is the major means of intercellular communication in metazoans (Hubbard & Miller 2007). Focusing on the effector side, receptor tyrosine kinase signaling starts with the binding of a growth factor to a receptor kinase, activating it. The second step in receptor tyrosine kinase signaling is phosphorylation of the primary substrates, often incorporated in the receptor tyrosine kinases themselves. The primary substrates are bound by cytoplasmic proteins through specialized protein domains, bringing them in close proximity to the activated receptor tyrosine kinase. Commonly, these proteins are then also phosphorylated by the receptor tyrosine kinase, or by a non-receptor tyrosine kinase recruited to a phosphotyrosine site in a primary substrate. Alternatively or additionally, they can serve as adapters for other proteins. Subsequently, the pathways branch out into apparently relatively clearly defined axes, like the PI3K-PKB-PTEN, the Ras-JNK or the JAK-STAT axis, and canonically end in nuclear factors, like JUN, FOS or MYC (Steelman et al. 2008). Misregulation of these pathways can lead to serious disease. For example, a mutation simulating a growth-inducing signal can cause cancer or an erroneous activation of immune receptor pathways can cause immunological disease (e.g. Müllauer et al. 2001).

1.2.2 Deregulation of phosphotyrosine signaling causes cancer

One of the most daunting challenges for the medical sciences today is cancer. In 2009, in the United States of America, the average lifetime risk of cancer is higher than 40% (Howlader et al. 2012). Cancer is the result of uncontrolled cell growth resulting from loss of regulation (Croce 2008). In cancer progression, the cells escape the control mechanisms mediated by apoptosis and cytotoxic T cells, acquire growth-inducing mutations and, finally, invade other tissues. Therefore, genes that are often found mutated in cancer comprise tumor suppressor genes, like RB1 and TP53, activating signaling components, like growth factors, kinases or transcription factors. Because cancers result from several mutations in the same cell and additional mutations can provide substantial growth advantages, there is another type of gene that is systematically altered in cancers. The third type of cancer genes is related to genetic stability. In principle, any alterations that makes acquiring further mutations easier are retained with high probability. (Vogelstein & Kinzler 2004) propose calling the growth activators "oncogenes" and the latter "stability genes". (Kinzler & Vogelstein 1997) use the term "caretakers" to describe the stability genes and "gatekeepers" for the other two groups. The distinction of cancer genes into caretakers and gatekeepers is very eidetic and calls attention to the association of specific mutations and cancer types. The distinction into oncogenes, tumor suppressor genes and stability genes is more precise, but requires additional insight. Moreover, the term "oncogene" can be misleading, because until the early 1990s it was commonly used to contrast the mutated, cancer-inducing forms of genes commonly found in tumor viruses from their wild type counterparts (called proto-oncogenes). In this classification, the most well-known genes related to phosphotyrosine signaling are growth activators. For example, many receptor tyrosine kinases, like KIT (Rubin et al. 2001), MET (Eder et al. 2009), RET (Santoro et al. 2004), AXL (Linger et al. 2010) or subunits of the epidermal (Nicholson et al. 2001) or fibroblast growth factor receptor (Turner & Grose 2010) have oncogenic potential when overactivated. Usually, these genes carry mutations that simulate growth factor binding. Non-receptor tyrosine kinases, especially Src family kinases and members of related families, on the other hand, are usually activated by mutations that neutralize intrinsic control mechanisms. SRC itself can be

Table 1: Associations between phosphotyrosine signaling genes and specific types of cancer indicate pathways important in particular cell types. Phosphotyrosine signaling genes associated with a specific type of cancer sorted alphabetically by Entrez Gene Symbol. For each gene, the Entrez GeneID, the type of cancer it is associated with and a reference discussing the association are listed.

GeneID	Symbol	Type of Cancer	Reference
25	ABL1	Chronic myelogenous leukemia	Mauro & Druker (2001)
558	AXL	Carcinoma	Li et al. (2009)
8412	BCAR3	Breast cancer	Wilson et al. (2013)
29760	BLNK	Pre-B cell acute lymphoblastic leukaemia	Jumaa et al. (2003)
695	BTK	Lymphoma/leukomia	Buggy & Elias (2012)
867	CBL	Myeloproliferative disorders	Naramura et al. (2011)
868	CBLB	Myeloproliferative disorders	Naramura et al. (2011)
23624	CBLC	Myeloproliferative disorders	Naramura et al. (2011)
1436	CSF1R	Carcinoma	Kacinski (1995)
1601	DAB2	Prostate/ovarian carcinoma	Fazili et al. (1999)
1956	EGFR	Non-small cell lung cancer	Rusch et al. (1993)
2322	FLT3	Acute myeloid leukemia	Warren et al. (2012)
3702	ITK	Peripheral T cell lymphomas	Streubel et al. (2006)
3932	LCK	Lymphoma/leukemia	Harr et al. (2010)
4058	LTK	Leukemia	Roll & Reuther (2012)
18212	NTRK2	Neuroblastoma	Nakagawara et al. (1994)
10019	SH2B3	Lymphoma/leukemia	Gery et al. (2009)
4068	SH2D1A	B cell neoplasms	Sandlund et al. (2013)
6464	SHC1	Mammary carcinoma	Northey et al. (2008)
53358	SHC3	Neuroblastoma	Miyake et al. (2009)
6714	SRC	Metastatic carcinoma	Frame (2002)

activated by mutations in its SH2 or SH3 domains or intramolecular SH2 or SH3 binding sites (Brown & Cooper 1996) and the common BCR-Abl translocation relieves regulation by an intramolecular SH3 domain interaction (Barilá & Superti-Furga 1998).

Interestingly, there is often a connection between certain cancer types and specific genes (Hanahan & Weinberg 2000). Table 1 shows a selection of tyrosine signaling genes with strong association to a specific type of cancer. This observation provides valuable insight into the mechanics of signaling in nascent cancer and, by extension, into phosphotyrosine signaling. In principle, all cancers arise from cells acquiring a set of mutations that allows an escape from growth regulation. Accordingly, one would expect that a specific genetic alteration has the same consequences for carcinogenesis irrespective of the cell or tissue type it occurs in. The fact that this is not so, shows that growth regulation differs between distinct cell types. Therefore, it seems likely that, at the very least, the degree to which the different components of the growth regulation interactome are relevant is cell type-specific. The same mutation can provide initial growth advantages and constitute the first step of carcinogenesis in one cell type while having little impact in another. As a consequence, that mutation is found more often in the cancer subtype related to the first cell type.
To sum up, phosphotyrosine signaling is highly relevant for cancer and the insights gleaned from cancer research help understanding the regulation of growth and development in normal, untransformed cells. Thus, a genome-scale map of phosphotyrosine-dependent protein-protein interactions will be medically important for a large part of the population.

1.2.3 Deregulation of phosphotyrosine signaling relates to immunological disease

Another medically relevant field with a strong connection to phosphotyrosine signaling is the regulation of the immune system. Many phosphotyrosine signaling proteins are named for their involvement in immunological processes, like B lymphocyte kinase (BLK), B cell linker (BLNK), Bruton agammaglobulinemia tyrosine kinase (BTK), macrophage colony stimulating factor I receptor (CSF1R), IL-2-inducible T cell kinase (ITK), lymphocyte-specific protein tyrosine kinase (LCK), lymphocyte cytosolic protein 2 (LCP2), macrophage stimulating 1 receptor (MST1R), or ζ-chain (TCR) associated protein kinase 70 kDa (ZAP70). Others have alternative names indicating immonological

functions, like mast cell immunoreceptor signal transducer (CLNK), B cell adapter molecule of 32 kDa (DAPP1), lymphocyte-specific adapter protein (SH2B3), SLP-76-associated protein of 130 kDa (FYB), signaling lymphocyte activation molecule-associated protein (SH2D1A), or T lymphocyte specific adaptor protein (SH2D2A).

Table 2 shows a selection of phosphotyrosine signaling genes with a clearly proven specific function in immunological cell function. Interestingly, there is a relatively large number of genes that have been shown to carry vital roles in several types of immune cells. This reflects the common evolutionary origin of the hematopoietic lineage. Even more interesting, however, is which functions concur. Mostly, the functions of a protein in different immunological cells are related. In other words, each of the proteins carries a biological meaning. What is striking about that, is that the semantic level is different. For example, SLA2 inhibits T cell receptor signaling in T cells (Pandey et al. 2002) and B cell receptor signaling in B cells (Holland et al. 2001). In contrast, HSH2D inhibits activation of T cells through the T cell receptor (Perchonock et al. 2006) but antagonizes BCR-induced apoptosis in B cells (Herrin et al. 2005) and Nck has been described as "critical for maturation of the immune synapse and for T cell activation" (Gil et al. 2002) while negatively regulating JNK activation and exerting a positive effect on apoptosis in response to BCR signaling in B cells (Mizuno et al. 2002). Therefore, HSH2D and Nck can be thought of as shifting immune response from T cell-mediated to B cell-mediated or vice versa, respectively, whereas SLA2 may be considered a general immune receptor antagonist. While these examples are very broad, others are more clearly defined. For instance, PLCG2 KO mice "have decreased mature B cells, a block in pro-B cell differentiation, and B1 B cell deficiency" and display "a loss of [...] mast cell FcεR function, and NK cell FcγRIII and 2B4 function" (Wang et al. 2000) and DAPP1 is involved in B cell response to TI-II antigens (Fournier et al. 2003) and is involved in Fcε RI-mediated mast cell degranulation (Hou et al. 2010). Thus, PLCG2 can be considered important for antibody-dependent cell-mediated cytotoxicity, an immune response that is typical for certain parasites, and DAPP1 for non-adaptive immune response to bacterial infections. BLK is "required for the development of wild-type numbers of IL-17-producing λδ T cells" (Laird et al. 2010) and regulates B cell tolerance (Hom

et al. 2008); LYN "plays a role in signal transduction for not only clonal expansion and terminal differentiation of peripheral B cells but also elimination of autoreactive B cells" (Nishizumi et al. 1995) and is required for in vivo phosphorylation of FcγRIIB in mast cells (Malbec et al. 1998). Hence, these LYN and BLK carry biological meanings similar to those of HSH2D and Nck in that they shift from cytotoxic to humeral immune response or the other way around, but on a much more specific level. In addition to rather specific immune-related phosphotyrosine signaling genes, there are many phosphotyrosine signaling genes better known from other contexts that are important in immune pathways, like PI3K (Deane & Fruman 2004), GRB2 (Gong et al. 2001) or PLCG2 (Wang et al. 2000). Conversely, a number of well-known immune genes is important in non-immune phosphotyrosine signaling contexts. For example, BLK is involved in insulin sectretion (Borowiec et al. 2009), SH3B3 is a negative regulator of PDGFR signaling (Gueller et al. 2011) and SH2D2A is required for SRC activation in VEGFR signaling (Sun et al. 2012). This wide range of immunological semantics covered by phosphotyrosine signaling genes brings up a number of questions. For example, the fact that some phosphotyrosine signaling proteins regulate different immune cells, like T cells and B cells, in the same manner while others regulate them in opposite ways suggests that the two proteins represent distinct semantic layers. It is tempting to speculate that the computation between the layers is a phosphotyrosine-related function. From an evolutionary perspective, we can ask why a system evolved in which the biological function of a number of proteins can be best understood on an organism level, i.e. why does protein function on the cellular level radically differ depending on cell type for some proteins? The apparent overlap between phosphotyrosine signaling genes with immune functions and phosphotyrosine signaling genes with non-immune functions that the pathways regulating immune cell action are embedded in a large regulatory network together with components common to other pathways making the adjustment of cell-specific signal integration the most frugal solution to the problems at hand. To validate or invalidate all of these suggestions, we need to consider the respective proteins in a network context.

In short, there is a strong connection of phosphotyrosine signaling and immunological control on many different levels. The general biological function

Table 2: Roles of phosphotyrosine genes in different kinds of immune cells are mostly similar or converse. Phosphotyrosine signaling genes with a clearly proven role in specific types of immune cells sorted alphabetically by Entrez Gene Symbol. For each gene, the Entrez GeneID, the function it has in T cells, B cells and other immune cells is listed with a reference discussing it, where known.

Entrez GeneID	Entrez Gene Symbol	Function in		
		T cells	B cells	other immune cells
640	BLK	development of IL-17-producing γδ T cells (Laird et al. 2010)	regulates B cell tolerance (Hom et al. 2008)	
29760	BLNK		transition from B220+CD43+ progenitor B to B220+CD43- precursor B cells (Pappu et al. 1999)	
695	BTK		important in B lymphocyte development, differentiation, and signaling (Mohamed et al. 2009)	
868	CBLB	regulates peripheral tolerance and anergy (Jeon et al. 2004)		
116449	CLNK			involved in FcεRI-mediated mast cell degranulation (Goitsuka et al. 2000)
1436	CSF1R			regulates the survival, proliferation, chemotaxis and activation of macrophages (Pixley & Stanley 2004)

Table 2: Roles of phosphotyrosine genes in different kinds of immune cells are mostly similar or converse. (continued)

Entrez GeneID	Entrez Gene Symbol	Function in		
		T cells	B cells	other immune cells
27071	DAPP1		response to TI-II antigens (Fournier et al. 2003)	involved in FcεRI-mediated mast cell degranulation (Hou et al. 2010)
2322	FLT3			activates and differetiates dendritic cells (Maraskovsky et al. 2000)
2534	FYN	NKT cell development (Gadue et al. 1999)		
84941	HSH2D	inhibits T cell activation (Perchonock et al. 2006)	antagonizes BCR-induced apoptosis (Herrin et al. 2005)	
3635	INPP5D	negatively regulates Th2 responses and T cell cytotoxicity (Tarasenko et al. 2007)	attenuates a proapoptotic signal initiated by FcγRIIB (Ono et al. 1997)	
3702	ITK	CD4+ T cell development (Liao & Littman 1995)		
3717	JAK2			DC-mediated innate immune response and sepsis (Zhong et al. 2010)
3932	LCK	thymocyte development (Molina et al. 1992)		
3937	LCP2	development of peripheral T cells (Clements et al. 1998)		FcεRI-mediated activation of mast cells (Pivniouk et al. 1999)

Table 2: Roles of phosphotyrosine genes in different kinds of immune cells are mostly similar or converse. (continued)

Entrez GeneID	Entrez Gene Symbol	Function in		
		T cells	B cells	other immune cells
4067	LYN		expansion and differentiation of peripheral B cells, elimination of autoreactive B cells (Nishizumi et al. 1995)	phosphorylation of FcγRIIB in mast cells (Malbec et al. 1998)
4690	NCK1	maturation of the immune synapse and T cell activation (Gil et al. 2002)	apoptosis in response to BCR signaling (Mizuno et al. 2002)	
8440	NCK2	maturation of the immune synapse and T cell activation (Gil et al. 2002)	apoptosis in response to BCR signaling (Mizuno et al. 2002)	
5336	PLCG2		maturation of B cells, differentiation of pro-B cell, B-1 B cell development (Wang et al. 2000)	mast cell FcεR function, NK cell FcγRIII and 2B4 function (Wang et al. 2000)
10603	SH2B2		negatively regulates B-1 B cell development (Iseki et al. 2004)	
10019	SH2B3		regulates pro-B cell proliferation (Takaki et al. 2000)	
4068	SH2D1A			CD244 function in NK cells (Tangye et al. 2000)
9047	SH2D2A	IL2 response and antigen-induced T cell death (Marti et al. 2005)		

Table 2: Roles of phosphotyrosine genes in different kinds of immune cells are mostly similar or converse. (continued)

Entrez GeneID	Entrez Gene Symbol	Function in		
		T cells	B cells	other immune cells
84174	SLA2	inhibits TCR signaling (Pandey et al. 2002)	inhibits BCR signaling (Holland et al. 2001)	
6772	STAT1			LPS-induced gene expression in macrophages (Ohmori & Hamilton 2001)
6778	STAT6	IL-4-mediated functions including Th2 helper T cell differentiation (Shimoda et al. 1996)	immunoglobulin class switching to IgE (Shimoda et al. 1996)	
6850	SYK	intraepithelial γδ T cells (Mallick-Wood et al. 1996)	differentiation of B-lineage cells (Cheng et al. 1995)	signaling from FcεRI in mast cells (Siraganian et al. 2002)
7409	VAV1	selection of CD4+CD8+ T cells (Turner et al. 1997)		
7535	ZAP70	development of CD4 and CD8 single-positive T cells (Negishi et al. 1995)		

associated with a phosphotyrosine signaling protein can be cellular or organismic. As a consequence, a number of questions regarding the processing of signals and the relation of immunological control to other phosphotyrosine signaling processes arise. Many of these questions are answered best from a network perspective. Therefore, a genome-scale map of phosphotyrosine-dependent protein-protein interactions can be a valuable foundation for future studies of immune function and immune-related disease, like autoimmune disorders, primary immune deficiency, or even allergies.

1.2.4 Tyrosine kinases, phosphatases and binders form a "toolkit"

One of the marked characteristics of biological processes, in contrast to, for example, chemical or physical processes, is that they are regulated (Keskin et al. 2004). As a general rule, the more complex the processes grow, the more the regulator and the regulans are separated. For example, in prokaryotes genes coding for proteins of related functions are often bundled together in operons, and transcription and translation are immediately linked, whereas, in eukaryotes, genes are almost always individually regulated on many levels and by many different mechanisms (Choudhuri 2004), like on the transcriptional level by transcription factors mediating RNA polymerase binding or epigenetic mechanisms modulating DNA accessibility, on the level of mRNA stability and translation by RNA interference, or on the level of protein stability by proteasomal degradation. Accordingly, there is a tendency towards higher modularity in more complex systems (Kashtan & Alon 2005).
There are qualitative evolutionary steps that are associated with drastic increases in complexity (Szathmáry & Smith 1995). One is the transition from prokaryotes to eukaryotes. Another one is the emergence of multicellularity. Multicellular organisms have a number of potential evolutionary advantages over unicellular ones. Mainly, they can achieve larger sizes and develop specialized cell types (Grosberg & Strathmann 2007). To take full advantage of this potential, however, several obstacles have to be overcome. Among these is the coordination of cells forming an organism to produce concerted and adequate behavior of the ensemble. In metazoans, this function is coupled, for

the most part, to protein tyrosine phosphorylation. It appears that, chronologically, phosphotyrosine signaling came first (Müller 2001), and it may even be a prerequisite for the evolution of complex metazoans (Hunter & Cooper 1985). Phosphotyrosine signaling seems well-suited for this purpose, primarily for two reasons: first, its evolutionary novelty provides a high potential for acquiring new functions without compromising old ones; second, it comes with a three component "toolkit" (Pincus et al. 2008) that allows fast and efficient regulation and transmission of signals. This toolkit consists of tyrosine kinase, SH2 or PTB and tyrosine phosphatase domains, fulfilling "writer", "reader" or "eraser" function, respectively. The evolutionary development of this toolkit merits special consideration, because the stabilizing evolutionary pressure is not obvious for the incomplete system. This is especially problematic for three component systems, because they imply two "preadaption" steps. In the case of phosphotyrosine signaling, it seems clear that the kinases came last. For example, the genome of *Saccharomyces cerevisiae* contains a (proto-)SH2 domain and several protein tyrosine phosphatases (Lim & Pawson 2010). Nevertheless, a minor degree of tyrosine phosphorylation has been observed (Gnad et al. 2009). Apparently, certain protein serine/threonine kinases are able to phosphorylate tyrosine residues. It has been suggested that these kinases might substitute for bona fide tyrosine kinases to some extent (Pincus et al. 2008). Indeed, there is evidence for phosphotyrosine-dependent regulation of the MAPK-related FUS3 kinase in *S. cerevisiae* by the dual-specificity kinase STE7 (Errede et al. 1993). On the other hand, the *S. cerevisiae* SH2 domain in the Spt6 protein has been shown to bind phosphoserine (Dengl et al. 2009). The first step is less clear. Although protein tyrosine phosphatase genes can be found even in prokaryotic genomes, their regulatory function, if they have any, is not understood (Kennelly 2002; Ramponi & Stefani 1997). It is therefore conceivable that these genes were acquired by lateral gene transfer mechanisms following the wide-spread tyrosine phosphorylation of proteins in eukaryotes, for example to allow metabolizing tyrosine-phosphorylated proteins or coopting the signaling potential (Kennelly & Potts 1996). The assembly of the complete "toolkit" was apparently quickly followed by an evolutionary burst, possibly even an increased rate of speciation (Pincus et al. 2008).

1.2.5 Protein tyrosine kinases are either membrane-spanning receptors or cytoplasmic proteins

Metazoan bona fide tyrosine kinases come in two flavors, receptor tyrosine kinases and non-receptor tyrosine kinases (Robinson et al. 2000).

1.2.5.1 Receptor tyrosine kinases are transmembrane proteins activated by extracellular ligand binding

Receptor tyrosine kinases (Ullrich & Schlessinger 1990) are transmembrane proteins that typically have an extracellular domain capable of specifically binding a small protein ligand, called growth factor, and inducing dimerization. The dimerization activates the intracellular kinase domain which then phosphorylates multiple tyrosine residues in the respective other monomer. Phosphorylation of tyrosine residues in the activation loop sustain kinase activation. Other residues form a scaffold for the binding of phosphotyrosine-binding proteins. The recruited proteins can carry enzymatic activities of their own or they may be substrates for the receptor tyrosine kinases or other recruited proteins or both. After being activated at the receptor kinase, the recruited proteins relay the signal into the cytoplasm and the nucleus.
This concept is realized in many growth factor/receptor pairs. In detail, there are often deviations from the principle. For example, the ERBB2 protein, one of the four subunits of the epidermal growth factor receptor is incapable of binding a ligand and relies on its dimerization partner (Klapper et al. 1999). Hence, in this case, the ligand does not induce dimerization but merely induces a conformational change leading to kinase activation. The insulin receptor has a ligand-binding domain that is encoded as another peptide chain and linked by a disulfide bridge and its scaffolding function is mostly assumed by the IRS1 and IRS2 genes. (Lee & Pilch 1994)

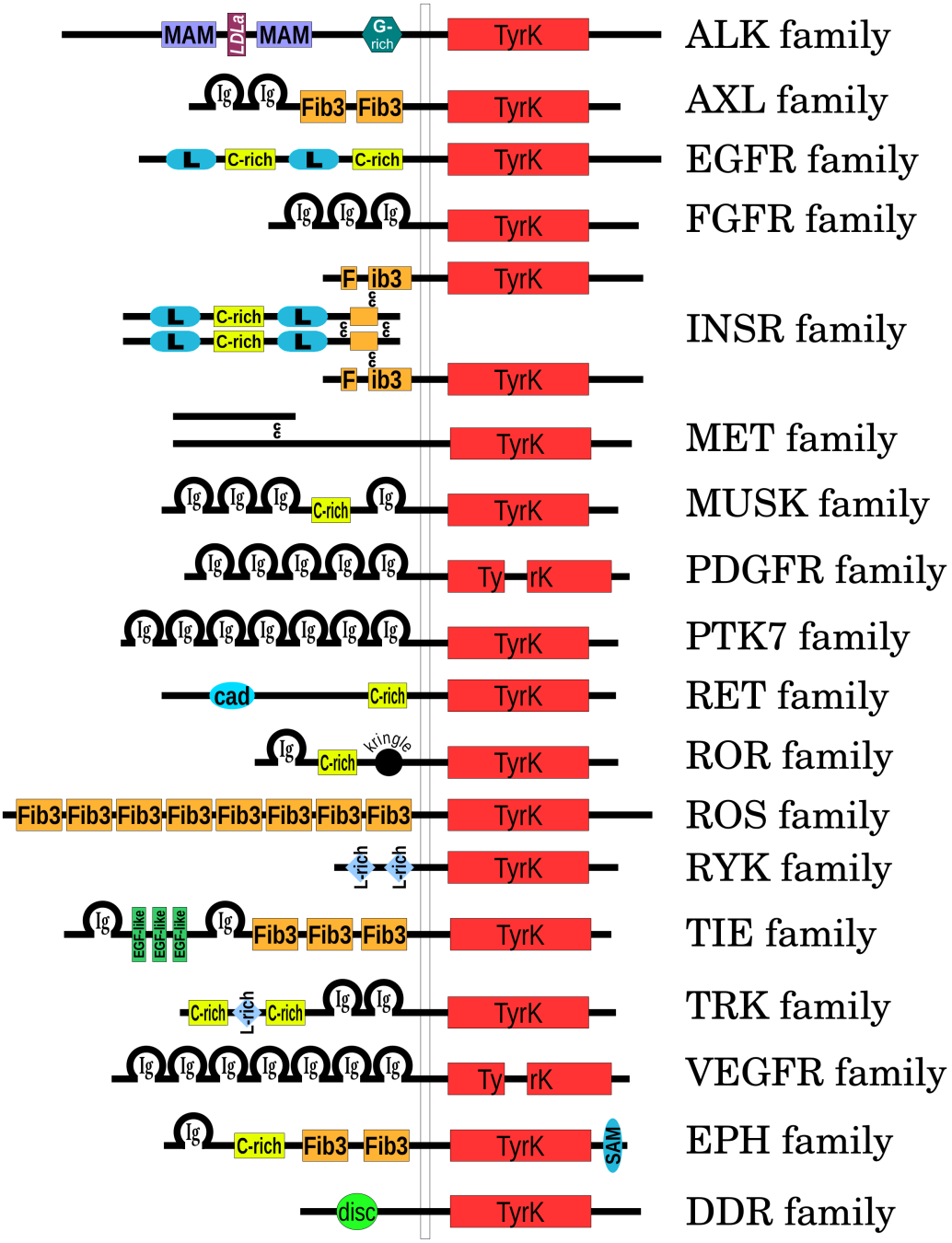
1.2.5.2 Receptor tyrosine kinases are classified into twenty families based on functional and structural properties

Upon systematic inspection, the receptor tyrosine kinases can be divided into twenty families (Robinson et al. 2000). The families differ mainly in the makeup of the extracellular part. Many receptor tyrosine kinase families are charac-

terized by immonoglobulin-like or fibronectin type 3 domains in this region, other domain types are specific for a single receptor tyrosine kinase family only (compare Figure 1).

Many growth factors can bind and activate several receptors and many receptors can bind and be activated by more than one growth factor (compare Figure 2). The growth factors are often most highly expressed in specific tissue or cell types. Similarly, the receptor tyrosine kinase complement of most cell types is highly specific. The apparent degree of growth factor/receptor promiscuity varies, as does the amount of research done on the different groups of growth factor/receptor pairs. For example, the EGFR family has been characterized in great detail and the differences binding and signaling properties of individual receptor dimers are known (Yarden & Sliwkowski 2001). Similarly, the many different FGF binding preferences have been mapped not only to the four FGFR genes, but even to specific exons (Zhang et al. 2006). Because of these differences, it is conceivable that there are relevant connections missing from our knowledge and that these connections preferentially exist between growth factors and receptors that are not obviously related. On the other hand, there are receptors for which no binding growth factors are known and these have received a lot of attention and, at least in some cases, seem to be fundamentally unable to bind any ligand. Nevertheless, these receptors can have signaling functions, like ERBB2 (Klapper et al. 1999). The same is true for catalytically inactive receptors. For example, the ERBB3 gene product is catalytically inactive (Guy et al. 1994) but it can heterodimerize with other EGFR gene products. Even the ERBB2/ERBB3 heterodimer is functional and physiologically relevant (Citri et al. 2003). The ERBB3 monomer binds the growth factor and changes its conformation allowing the catalytically active ERBB2 monomer to phosphorylate it (Graus-Porta et al. 1997). An even more peculiar example is RYK, which is also catalytically inactive (Katso et al. 1999). It has been shown to function as a receptor for Wnt pathway ligands, like

Figure 1: Receptor tyrosine kinase families are defined by domain composition of the extracellular region. Schematic depicting the domain composition of receptor tyrosine kinase families. For each family of receptor tyrosine kinase implicated in signaling, the protein domain composition is drawn approximately to scale. Kinases are sorted alphabetically and aligned by their



transmembrane region, indicated by a continuous vertical block. C-rich = cysteine-rich domain; cad = cadherin domain; CC = disulfide bond; disc = Discoidin domain; EGF-like = EGF-like domain; Fib3 = Fibronectin type III domain; G-rich = glycine-rich domain; Ig = immunoglobulin domain; kringle = kringle domain; L = L domain; L-rich = leucine-rich domain; LDLa = low density lipoprotein class A domain; MAM = MAM domain; SAM = sterile alpha motif domain; TyrK = tyrosine kinase domain. Adapted from (Hubbard & Till 2000) and modified.

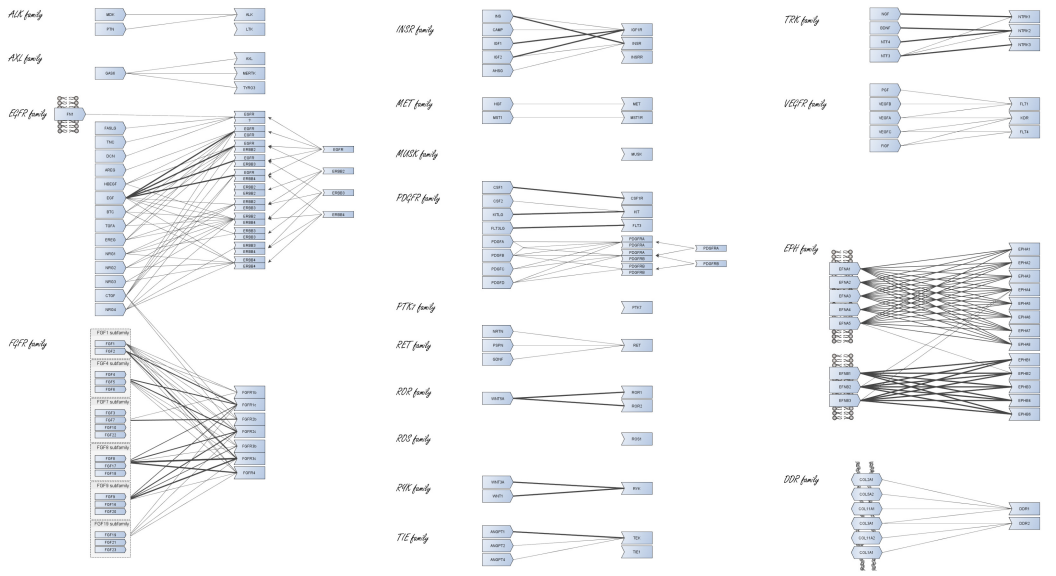


Figure 2: Growth factors and receptor tyrosine kinases display diverse degrees of promiscuity and binding patterns seem to be governed by receptor tyrosine kinase family membership. Schematic depicting the binding preferences of growth factors and growth factor receptor tyrosine kinases. Growth factor receptor tyrosine kinases are arranged alphabetically by family, growth factors are placed close to the binding receptors. Convex rectangle-like polygons symbolize growth factors; concave rectangle-like polygons symbolize growth factor receptors. Lines connecting growth factors and receptors indicate ability to bind each other. Where known, higher affinity binding is reflected by thicker lines. Transmembrane proteins are drawn before a cartoon representation of a lipid bilayer; proteins integral to the extracellular matrix are drawn before a cartoon representation of tangled helical protein strings. Because the EPH and DDR families of receptor tyrosine kinases bind transmembrane or extracellular matrix ligands, respectively, they are shown separately in the lower right corner.

WNT1 and WNT3A (Lu et al. 2004). Interestingly, it also associates with and attenuates the signaling of Ephrin receptors (Trivier & Ganesan 2002; Halford et al. 2000).

1.2.5.2.1 A small number of proteins seems to be central to most receptor tyrosine kinase pathways

In short, the complement of human receptor tyrosine kinases embodies a multifarious reiteration of an efficient signaling paradigm. In addition to this large variety in growth factor binding preference and signaling mechanisms,

the receptor tyrosine kinases also appear to fulfill many different, quite specific biological roles. This is reflected in the corresponding gene knock-out phenotypes (Table 3). Interestingly, receptor tyrosine kinases signal through

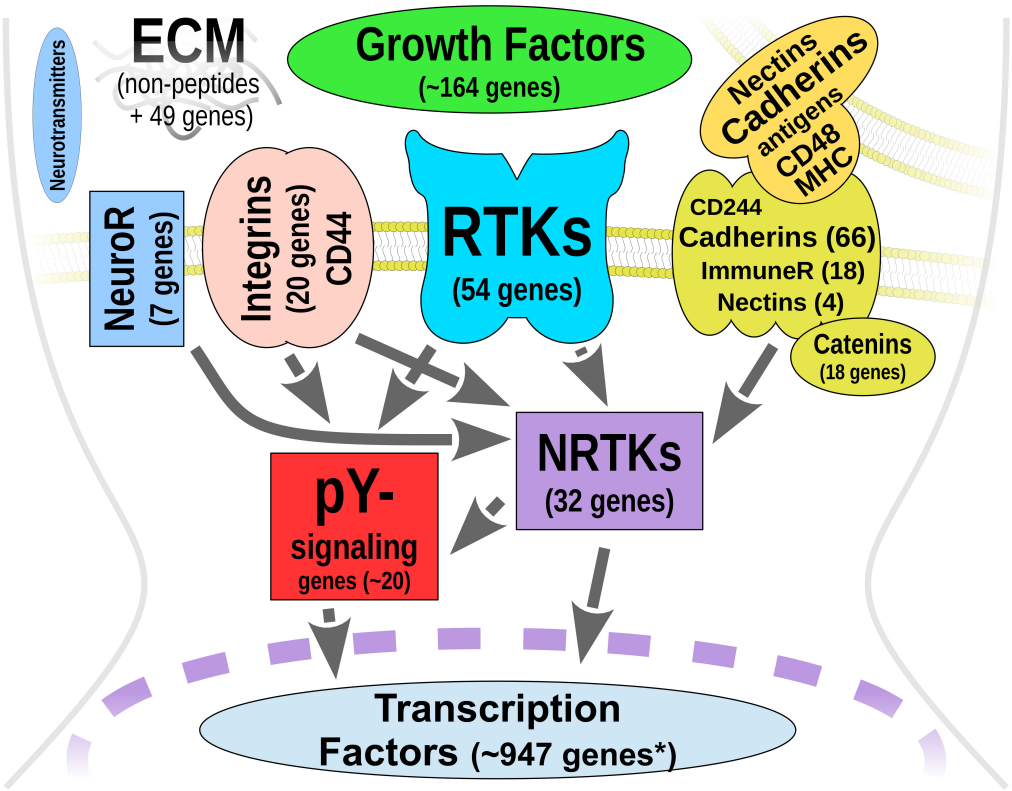


Figure 3: A small set of phosphotyrosine signaling components is central to the "hourglass" structure of growth factor receptor signaling. Schematic depicting compound signaling pathways related to cell-cell communication. Extracellular signals presenting in the form of neurotransmitters, components of the extracellular matrix (ECM), soluble growth factors and proteins integral to the membranes of neighboring cells, like nectins, cadherins, antigens presented, CD48 or the main histocompatibility complex (MHC), far outnumber their respective receptors. These receptors, comprising of neurotransmitter receptors (NeuroR), integrins, receptor tyrosine kinases (RTKs), CD244, cadherins, immune cell receptors (ImmuneR), nectins and catenins, in turn, relay the signals through an even smaller number of cytoplasmic non-receptor tyrosine kinases (NRTKs) and phosphotyrosine-recognizing domain-containing signaling genes (pY-signaling genes) to achieve very specific transcription profiles appropriate to the extracellular signals.[a]

[a]Note that the number of transcription factors provided is an estimate based on the number of genes annotated with the term GO:0003700 "sequence-specific DNA binding transcription factor activity" in the Gene Ontology project.

Table 3: Receptor tyrosine kinase mouse knock-out phenotypes are often very severe. Overview of obvious phenotypic effects of receptor tyrosine kinase gene knock-outs in *M. musculus*. Genes are sorted alphabetically by gene family and NCBI Symbols. Knock-out effects are described in terms of easily observed differences to wild type animals and effects on life expectancy, in particular. For each knock-out, a reference discussing its effects is provided.

Family	Gene Symbol(s)	Obvious Phenotype	Life Expectancy Effect	Reference
ALK	ALK	Alterations in adult brain function	-	Bilsland et al. (2008)
	LTK	(No information available)		
AXL	AXL	-	-	Lu et al. (1999)
	MERTK	Retinal degradation	-	Duncan et al. (2003)
	TYRO3	-	-	Lu et al. (1999)
	AXL, MERTK, TYRO3	Male sterility, blindness	-	Lu et al. (1999)
DDR	DDR1	Reduced size and body weight	-	Vogel et al. (2001)
	DDR2	Dwarfism, impaired wound healing	-	Labrador et al. (2001)
EGFR	EGFR	Abnormalities in skin, kidney, brain, liver, and gastrointestinal tract	Lethal[1]	Threadgill et al. (1995)
	ERBB2	Lack of cardiac trabuulae; compromised motor nerve development	Lethal (E11)	Lee et al. (1995)
	ERBB3	Severe neuropathies	Lethal (E12.5)	Riethmacher et al. (1997)
	ERBB4	Aborted development of the heart muscle; abnormal hindbrain development	Lethal (E10.5)	Gassmann et al. (1995)
EPH	EPHA1	Abnormal tail development; imperforate uterovaginal development	-	Duffy et al. (2008)
	EPHA2	-	-	Chen et al. (1996)
	EPHA3	Cardiac abnormalities	75% Lethal (P1)	Stephen et al. (2007)
	EPHA4	Peroneal muscular atrophy; absence of peroneal nerve	-	Helmbacher et al. (2000)
	EPHA5	-	-	Feldheim et al. (2004)
	EPHA6	-	-	Savelieva et al. (2008)
	EPHA7	-	-	Rashid et al. (2005)

[1] exact time depends on genetic background

Table 3: Receptor tyrosine kinase mouse knock-out phenotypes are often very severe. (continued)

Family	Gene Symbol(s)	Obvious Phenotype	Life Expectancy Effect	Reference
EPH	EPHA8	-	-	Park et al. (1997)
	EPHB1	-	-	Williams et al. (2003)
	EPHB2	-	-	Orioli et al. (1996)
	EPHB3	-	-	Orioli et al. (1996)
	EPHB4	Defective cardiovascular development	Lethal (E10.5)	Gerety et al. (1999)
	EPHB5		(Chicken only)	
	EPHB6	-	-	Shimoyama et al. (2002)
	EPHB2, EPHB3	Secundary palate closure failure	75% Lethal (P1)	Orioli et al. (1996)
FGFR	FGFR1	-	Lethal (<E10.5)	Deng et al. (1994)
	FGFR2	-	Lethal (<E8.5)	Arman et al. (1998)
	FGFR3	Skeletal overgrowth; inner ear defects	Reduced lifespan	Colvin et al. (1996)
	FGFR3, FGFR4	Growth retardation, sickly appearance, failure of secondary septation in the lungs	Severely reduced lifespan	Weinstein et al. (1998)
	FGFR4	-	-	Weinstein et al. (1998)
INSR	INSR	Growth retardation, pronounced hyperglycemia	Lethal (<P7)[2]	Joshi et al. (1996)
	INSRR	-	-	Kitamura et al. (2001)
	IGF1R	Respiratory failure, severe growth retardation	Lethal (P1)	Liu et al. (1993)
MET	MET		Lethal (E12.5)	Bladt et al. (1995)
	MST1R	-	-	Waltz et al. (2001)
MUSK	MUSK	-	Lethal (P1)	DeChiara et al. (1996)
PDGFR	PDGFRA	Multiple neuronal, skeletal and cardiac abnormalities	Lethal (<E16)	Soriano (1997)
	PDGFRB	Hemorrhagic, thrombocytopenic, severely anemic	Lethal (P1)	Soriano (1994)
	KIT	-	Lethal (<P1)	Nocka et al. (1990)

[2]Note that INSR deficiency occurs naturally and is apparently not lethal in human (Wertheimer et al. 1993).

Table 3: Receptor tyrosine kinase mouse knock-out phenotypes are often very severe. (continued)

Family	Gene Symbol(s)	Obvious Phenotype	Life Expectancy Effect	Reference
PDGFR	FLT3	-	-	Mackarehtschian et al. (1995)
	CSF1R	Toothlessness, truncated limbs, domed skull, low body weight, low growth rate, deafness	-	Dai et al. (2002)
PTK7	PTK7	-	-	Lee et al. (2012)
RET	RET	Renal agenesis or severe dysgenesis; disrupted enteric neurogenesis	Lethal (P1)	Schuchardt et al. (1994)
ROR	ROR1	-	Lethal (P1)	Nomi et al. (2001)
	ROR2	Dwarfism, severe cyanosis	Lethal (P1)	Takeuchi et al. (2000)
ROS	ROS	Male sterility	-	Sonnenberg-Riethmacher et al. (1996)
RYK	RYK	Shortened snout and limbs, completely cleft secondary palate	Lethal (P1)	Halford et al. (2000)
TIE	TIE	Severe edema, ruptured microvasculature	Lethal (E14)	Puri et al. (1995)
	TEK	Growth retardation, malformation of vascular network	Lethal (E9.5)	Dumont et al. (1994)
TRK	NTRK1	Severe neuropathies	Lethal (<P55)	Smeyne et al. (1994)
	NTRK2	-	Lethal (<P1)	Klein et al. (1993)
	NTRK3	-	Lethal (<P21) ("most")	Klein et al. (1994)
VEGFR	KDR	No organized blood vessels	Lethal (E9)	Shalaby et al. (1995)
	FLT1	Abnormal vascular channels	Lethal (E8.5)	Fong et al. (1995)
	FLT4	Abnormally organized large blood vessels	Lethal (E9.5)	Dumont et al. (1998)

a relatively small number of cytoplasmic proteins, like non-receptor tyrosine kinases, GTPases, phosphoinositol and protein kinases and phosphatases and adaptors, but ultimately effect a wide range of very specific responses (Figure 3). This phenomenon is pictorially referred to as "hourglass" or "bow-tie" structure (Oda et al. 2005). This structure has been argued to provide heightened robustness (Ma & Zeng 2003; Kitano 2004), which seems peculiar given that more than half of the receptor tyrosine kinase genes have a lethal phenotype when knocked out in *Mus musculus*, but actually just underlines the importance of phosphotyrosine signaling. More intriguing, however, is the question how the signal specificity is retained while passing through the "neck" in the "hourglass" (or the "knot" in the "bow-tie") structure. Towards this end, the intracellular phosphotyrosine signaling network has to be scrutinized.

1.2.5.3 Non-receptor tyrosine kinases are intracellular effectors of protein phosphotyrosine signaling

The other kind of protein tyrosine kinase found in higher eukaryotes is the group of non-receptor tyrosine kinases (Figure 4). These kinases are typically cytoplasmatic and become activated upon being covalently modified or binding an activated protein (Neet & Hunter 1996). Thus, they can function downstream of the receptor tyrosine kinases in growth factor receptor pathways or they can react to other signals, like chemokine receptor or integrin activation. In the case of growth factor receptor signaling, the non-receptor tyrosine kinases typically bind to sites that only become available after growth factor binding, like a previously non-phosphorylated tyrosine residue on the receptor tyrosine kinase itself or an associated adaptor protein. The most prominent non-receptor tyrosine kinase family is the Src kinase family. Many Src kinase family members have initially been found in tumor viruses (Gschwind et al. 2004). The reason for this lies in the easily activatable structure and high potential for promoting cell growth. Like most non-receptor tyrosine kinases, the Src family kinases have a C-terminal kinase domain. The regulatory N-terminal part consists of a "unique" signal peptide sequence, an SH3 and an SH2 domain. The signal peptide often contains a site for covalent modification with a fatty acid residue, that directs the protein to the plasma membrane and activates it. The SH3 and SH2 domains can fold back to bind intramolecular sites, rendering

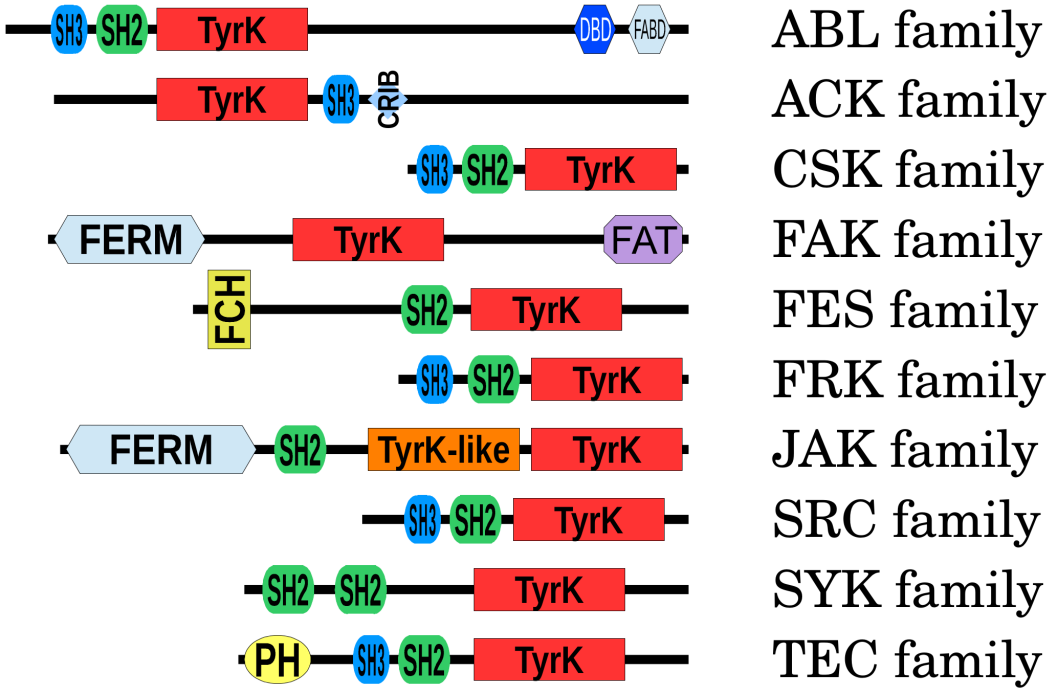


Figure 4: Non-receptor tyrosine kinase families contain protein-protein interaction domains. Schematic depicting the domain composition of non-receptor tyrosine kinase families. For each family of non-receptor tyrosine kinase the protein domain composition is drawn approximately to scale. Kinases are sorted alphabetically and aligned by their carboxy-terminus. CRIB = Cdc42/Rac interactive binding domain; DBD = DNA-binding domain; FABD = F-actin-binding domain; FAT = Focal adhesion targeting domain; FCH = Fes/CIP4 homology domain; FERM = FERM domain; PH = Pleckstrin homology domain; SH2 = Src homology 2 domain; SH3 = Src homology 3 domain; TyrK = Tyrosine kinase domain; TyrK-like = Tyrosine kinase-like domain (catalytically inactive). Adapted from Blume-Jensen & Hunter (2001) and modified to resemble the appearance in Hubbard & Till (2000).

the kinase domain catalytically inactive (Harrison 2003). To do so, the SH2 domain requires a C-terminal tyrosine residue to be phosphorylated (Y527 in SRC). Canonically, this residue is exclusively phosphorylated by the non-receptor tyrosine kinase CSK. Consequently, these kinases can be activated in several different ways. Besides the fatty acid modification, the intramolecular domain interactions can be released by providing competing binding sites or by regulating the phosphorylation status of the kinase itself. In addition to the N-terminal inhibitory tyrosine residue, there is an activating tyrosine residue in the activation loop within the kinase domain. Possibilities for activity-modu-

lating mutations are equally manifold. Point mutations rendering fatty acid modification obligatory or resembling a phosphotyrosine residue in the activation loop have an activating effect, while disruption of fatty acid modification or a tyrosine-to-phenylalanine mutation in the activation loop have an inactivating effect. Moreover, and more relevant for tumor viruses, both, N- and C-terminal truncation of the kinase gene, can release intramolecular inhibition and activate Src family kinases to a degree that brings about transforming potential. Under healthy physiological conditions, however, activation usually requires binding to a protein that has limited motility or undergoes transition to its activated form only at very specific subcellular localizations. As a result, non-receptor tyrosine kinase domain specificity is often low, because activation is linked to binding and localization. Another consequence of this property is that non-receptor tyrosine kinases are often active when transferred to an unnatural environment, since the inhibiting factors, like CSK, are absent.
In short, non-receptor tyrosine kinases are important players in the context of phosphotyrosine signaling, are highly relevant for cancer development, are often also phosphotyrosine binders and are usually "naturally active", making them very suitable for application out of the native cellular environment.

1.2.6 Phosphotyrosine-mediated protein-protein interactions are mediated by specialized protein domains

Protein tyrosine phosphorylation is a post-translational modification that provides a mechanism for highly efficient dynamic regulation of protein-protein interactions, both in terms of low latency and slim energy costs. Due to the small size of the phospho group, the recognizing domains have to be optimized for differentiating very similar structures. At the same time, phosphotyrosine-recognizing proteins have to be specific for their binding partners to allow the construction of complex regulatory networks. This specificity can be conferred by the immediate surroundings of the phosphorylation site or by additional, allosteric sites, or both. Early peptide binding experiments assume that binding specificity stems from linear phosphotyrosine peptide motifs and the structure of the phosphotyrosine-recognizing domain, alone. Apparently, these peptide motif/binding domain specificities explain most of the biologically relevant phos-

photyrosine-dependent protein-protein interactions. Nevertheless, it is possible that the peptide binding experiments gave direction to the field of phosphotyrosine-dependent signaling and, therefore, possible binding events not fitting the prevalent paradigm may just have been discarded without being examined. This notion is supported by recent findings indicating that many phosphorylated tyrosine residues in proteins do not fit the peptide motif knoledge (Olsen et al. 2006) and that multi-site tyrosine phosphorylation of proteins and multi-site phosphotyrosine-recognition by the same interaction partner is the rule rather than the exception (Schulze et al. 2005; Hanke & Mann 2009). Similarly, several proteins contain more than one phosphotyrosine-recognizing domain and the normal function of these arrangements seems binding to several sites on the same interaction partner (O'Brien et al. 2000).
Phosphotyrosine recognition is mediated mainly by two types of domains: Src Homology 2 (SH2) domains and PhosphoTyrosine Binding (PTB) domains. Additionally, a small fraction of Pleckstrin Homology (PH) domains has been shown to be able to recognize phosphotyrosine moieties. Finally, there are even occasional reports of phosphotyrosine recognition by other domains.

1.2.7 SH2 domains are the primary mediators of phosphotyrosine-dependent protein interactions

The protein domain most closely linked to phosphotyrosine recognition is the Src Homology 2 (SH2) domain. This domain was named for the non-receptor tyrosine kinase SRC and has first received attention, because it was conserved among cytoplasmic, but not membrane-spanning, tyrosine kinases and was able to influence kinase activity in a chimeric viral FES-related kinase gene (Sadowski et al. 1986). The SH2 domain is a conserved region that is found in 121 instances in 111 human genes (Liu et al. 2011a), nearly all of which are known to recognize phosphotyrosine residues in proteins.

1.2.7.1 SH2 domain structure is highly conserved and contains an arginine residue critical for phosphotyrosine binding

The canonical SH2 domain fold consists of two α-helices and seven β-strands arranged in the sequence β-α-β-β-β-β-α-β (Figure 5). The parts are labeled

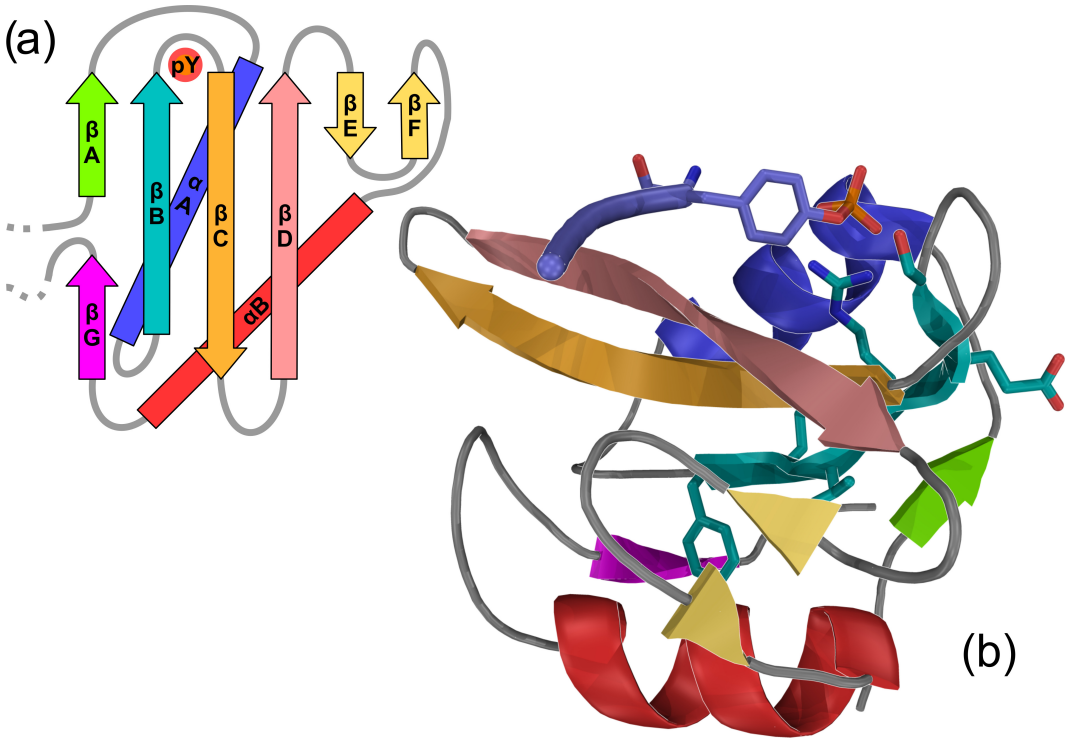


Figure 5: Src homology 2 (SH2) domains share a common fold with a central, highly conserved motif responsible for phosphotyrosine recognition. Schematic depicting the fold of a prototypic Src homology 2 (SH2) domain as 2-dimensional (a) and 3-dimensional structure model (b). Secondary structural elements are labeled in the 2-dimensional structure model (a) and color-coded identically in the two models. In the 3-dimensional structure model (b), the conserved FLVRES-motif and the recognized phosphotyrosine residue (pY) are shown as sticks, the rest of the binding peptide (light blue) and backbone of the SH2 domain are shown as cartoons. The 3-dimensional structure model (b) is based on the Protein Data Bank (PDB) entry 1SHA, a crystal structure model of the v-Src SH2 domain (Waksman et al. 1992).

by a secondary structure denominator and a consecutive capital Roman letter. Loops and turns are addressed by two Roman letters, corresponding to the connected secondary structure elements. The most conserved part of the SH2 fold is made up by the two α-helices flanking a central anti-paralell β-sheet made up by three β-strands, βB, βC and βD. The interaction interface is formed by βD, the N-terminal part of αA and less conserved features specific to the individual SH2 domain gene. βE and βF form an extension of the anti-paralell β-sheet, annealing to the C-terminal part of βD. βD can be split into two parts, referred to as βD and βD′. The phosphotyrosine residue binding cleft is located between

αA and the β-sheet; the C-terminal extension of the tyrosine-phosphorylated peptide or protein proceeds perpendicular to the β-sheet towards αB. In this region, there are often gene-specific features contributing to binding specificity, like additional loops or α-helices. The N- and C-termini of the SH2 domain lie at the side opposite to the binding surface and are canonically connected by βA and βG. Thus, βA and βG are usually very short and often hard to recognize as β-strands, often appearing highly bent with one end approaching a parallel β-sheet with βB.

Accordingly, the binding surface is usually on the most exposed side of the SH2 domain. The recognition of the phosphotyrosine residue itself is ensured by several contacts. The FLVRES-motif, the most highly conserved peptide motif in the SH2 domain, is central to βB. The first three residues position βB in relation to the other features. The arginine residue makes two hydrogen bonds to two terminal oxygen atoms in the phospho group with its primary amino groups. It is often stabilized by a hydrogen bond between a serine or threonine residue in βC and the secondary amine connecting the guanidinium group. According to recent knowledge, this residue is indispensable for phosphotyrosine recognition by SH2 domains (Mayer et al. 1992). The glutamate residue in the FLVRES-motif is not directly involved in phosphotyrosine recognition and actually points away from the binding surface. Nevertheless, it is highly conserved and is often involved in several hydrogen bonds with αA and αB or additional structural elements, like loops. The FLVRES serine residue makes a hydrogen bond to the oxygen atom connecting the phospho group to the phenyl group. The residue immediately C-terminal to the FLVRES-motif makes a hydrogen bond to the phospho group with its backbone amine. The next residue is often a serine or threonine that makes another hydrogen bond to the last oxygen atom in the phospho group. The phenyl ring of the phosphotyrosine residue is recognized by two basic amino acid residues forming amino-aromatic interactions. There is a conserved arginine residue at the N-terminal end of αA, that typically also makes a hydrogen bond to the phospho group. The second one is either a lysine or an arginine residue in βD. This residue is often preceded by an aromatic residue, that contributes to the binding surface. Before that, there is a highly conserved histidine residue that contacts and positions the arginine residues in αA and the FLVRES arginine residue.

Overall, the SH2 domain is well-conserved structurally and in terms of actual amino acid sequence. This is especially true for the residues involved in phosphotyrosine recognition and the residues forming the fold-determining secondary structures.

1.2.7.2 SH2 proteins often contain other domains defining their functions

The complement of SH2 domain-containing proteins is commonly divided into groups according to biological function, as biological function and domain composition are strongly related (Liu et al. 2006). Some SH2 proteins have catalytic domains. For example, many non-receptor tyrosine kinases are SH2 proteins (Neet & Hunter 1996). Similarly, there are protein tyrosine phosphatases with SH2 domains (Neel et al. 2003). Interestingly, there are apparently no protein serine or threonine kinases or phosphatases containing SH2 domains. Nevertheless, a number of SH2 proteins carry domains related to phosphoinositol signaling (Carpenter & Cantley 1996), connecting phosphotyrosine signaling to phosphoserine/-threonine signaling. Specifically, PLCG1 and PLCG2 cleave the phospholipid phosphatidylinositol 4,5-bisphosphate (PIP2) into diacyl glycerol (DAG) and inositol 1,4,5-trisphosphate (IP3). DAG activates the serine/threonine kinase protein kinase C (PKC). IP3 causes a release of intracellular Ca2+ and activates Ca2+/calmodulin-dependent protein kinases and PKC. The regulatory subunits of PI3K are also SH2 proteins, the catalytic subunits phosphorylate PIP2 producing phosphatidylinositol 3,4,5-trisphosphate (PIP3). PIP3 activates PDK1 and AKT kinases, providing another link to phosphoserine/-threonine signaling. Finally, two of the phosphatases that inactive PIP3, INPP5D and INPPL1, are also SH2 proteins. Thus, while there is no SH2 protein with a domain related to phosphoserine/-threonine signaling, there are many indirect links between the two. Other catalytic domains found in SH2 domains are more generic with regard to biological process. STAT proteins have DNA binding domains identifying them as transcription factors, the Cbl family proteins have RING and UBA domains, because they are E3 ubiquitin ligases. SUPT6H has a chromatin remodelling domain and many SH2 proteins have domains related to G-protein function. CHN1 and CHN2 have RhoGAP domains, VAV1, VAV2 and VAV3 have RhoGEF domains; RASA1 has a RasGAP

domain, SH2D3A, SH2D3C and BCAR3 have RasGEF domains, and RIN1, RIN2 and RIN3 have Ras association domains. In addition to SH2 proteins with domains related to other functions, some SH2 proteins have functions that are not associated to a specific protein domain, like the tensin proteins, all of which have cytoskeletal regulatory functions, or have binding domains only, making their cellular function less obvious. By virtue of their many binding domains and binding sites, proteins in the latter group are often able to bind many interaction partners. Their function is therefore usually thought of as bringing several other proteins together. For this reason, they are often referred to as adaptor or scaffold proteins. Other protein interaction domains that are often found in SH2 proteins are PTB, PH and SH3 domains, as well as SOCS boxes.

To sum up, SH2 domains are often paired with other domains and fulfill many different molecular functions related to several biological processes. Additionally, some SH2 proteins seem to function as adaptor proteins, connecting tyrosine-phosphorylated proteins to other proteins. Many SH2 proteins are also tyrosine-phosphorylated and have been shown to interact with other SH2 proteins. As a result, much of the phosphotyrosine function knowledge is related. On the other hand, due to the knowledge-driven nature of low throughput experiments investigating specific biological function, phosphotyrosine-dependent interaction not fitting the prevalent paradigms may have been simply overlooked. Therefore, the expected network resulting from an unbiased, genome-scale screen for phosphotyrosine-dependent protein-protein interactions is a, for the most part, connected network, that may or may not show a modular or hierarchical structure, but no systematic study involving full-length proteins has been produced so far.

1.2.8 Many PTB domains bind phosphotyrosine residues

The second major phosphotyrosine-recognizing domain is the phosphotyrosine binding (PTB) domain. The first description of a PTB domain was of the SHC1 PTB domain (Blaikie et al. 1994). In an effort to find new binders of tyrosine-phosphorylated EGFR (Blaikie et al. 1994) randomly primed cDNA from mouse fibroblasts was tested for binding to several human EGFR constructs. They isolated a clone that was highly similar to the human SHC1 gene, but did

not contain the SH2 domain, and corroborated this result by showing that the homologous region in the human gene also binds EGFR in a tyrosine phosphorylation-dependent fashion. Another such non-SH2 phosphotyrosine-recognizing domain was found in IRS1 (Wolf et al. 1995) and, in spite of low primary sequence similarity, adopts a similar fold (Zhou et al. 1996). This domain represents another type of PTB domain. Subsequently, PTB domains were classified as either "Shc-like" or "IRS-like". Despite the suggestive name, the PTB domain is actually less strongly implicated in phosphotyrosine recognition than SH2 domains. In fact, (Uhlik et al. 2005) divide the PTB domains into three groups and classify the largest group, comprising 75% of all PTB domains, as phosphorylation-independent protein tyrosine binding. The PTB domain groups heterogeneity is reflected in the fact that the members of the newly proposed group of (non-phosphotyrosine-binding) "Dab-like" PTB domains is more similar to the (phosphotyrosine-binding) "Shc-like" PTB domains than the (also phosphotyrosine-binding) "IRS-like" PTB domains are to each other.

1.2.8.1 PTB domains are structurally similar to PH domains

PTB domains represent a subgroup of domains with a pleckstrin homology (PH) superfold, and are sometimes referred to as PH domains themselves (Blomberg et al. 1999). As such, they are characterized by two antiparalell β-sheets forming a perpendicular sandwich structure and an amphipatic C-terminal α-helix contacting both β-sheets at the edge of the sandwich structure (Figure 6). Peptides are bound at an interface between the C-terminal helix and one of the β-sheets. Both β-sheets consist of four strands with one strand being part of both sheets, displaying a strong bend in the middle. Labeling the parts by a secondary structure denominater and a consecutive capital Roman letter the first sheet is formed by the N-terminal part of βA, βB, βC and βD. The second one is formed by βE, βF, βG and the C-terminal part of βA. αA connects βA and βB. αB is the conserved C-terminal helix and contacts the β-sheets along βE and the N-terminal half of βA in an "antiparalell" manner. The binding surface is formed by αB and βE. The bound peptide aligns to the second β-sheet, forming an antiparalell extension. This is referred to as antiparallel β-sheet augmentation (Harrison 1996). Deviations from this basic scheme are relatively seldom and slight. For example, the PTB domains of SHC1, DAB1

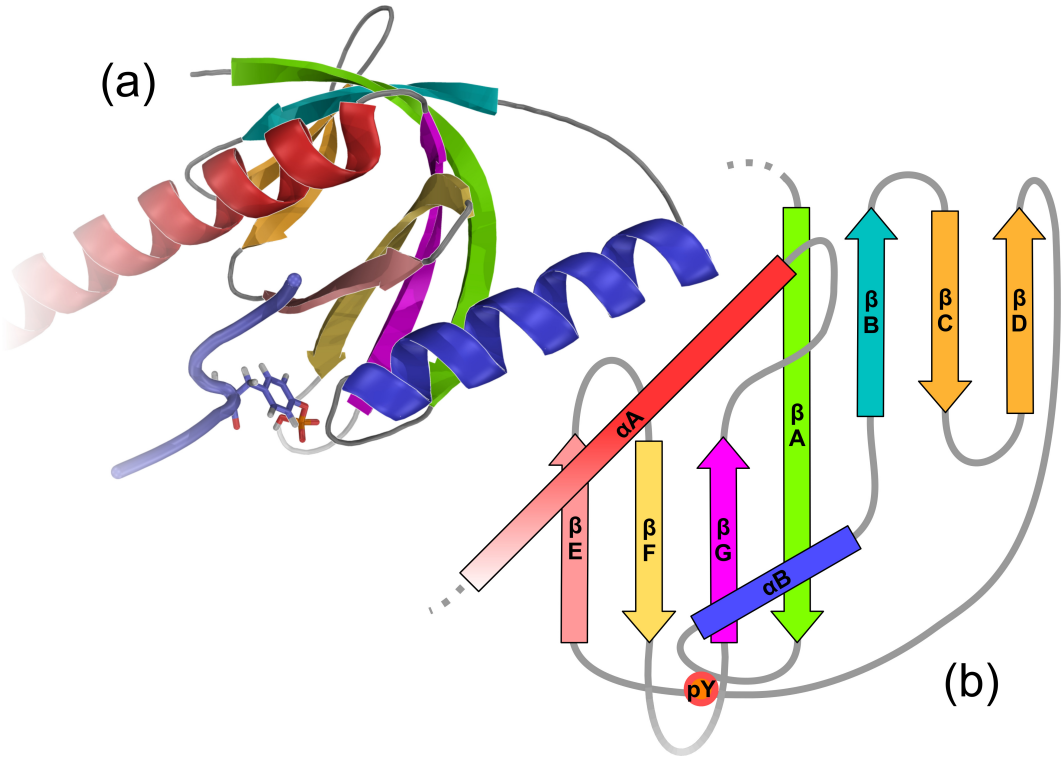


Figure 6: Phosphotyrosine binding (PTB) domains share a fold with Pleckstrin homology (PH) domains. Schematic depicting the fold of a prototypic Phosphotyrosine binding (PTB) domain as 3-dimensional (a) and 2-dimensional structure models. Secondary structural elements are labeled in the 2-dimensional structure model (b) and color-coded identically in the two models. In the 3-dimensional structure model (a), the amino-terminal α-helix is only partly shown. The recognized phosphotyrosine residue (pY) is shown as sticks, the rest of the binding peptide (light blue) and backbone of the PTB domain are shown as cartoons. The 3-dimensional structure model (a) is based on the Protein Data Bank (PDB) entry 2ELA, a crystal structure model of the APPL1 PTB domain (Li et al. 2007), the recognized peptide was modeled based on the Protein Data Bank (PDB) entry 1SHC, a crystal structure model of the SHC1 PTB domain (Zhou et al. 1995).

and NUMB have an extra α-helix N-terminal to the first β-strand (Farooq et al. 2003; Yun et al. 2003; Zwahlen et al. 2000); in the IRS1 PTB domain βA does not partake in the second β-sheet (Zhou et al. 1996); in the SHC1 PTB domain, βA is split into two β-strands and αA is followed by a short additional β-strand that extends the second anti-paralell β-sheet next to the second part of βA (Farooq et al. 2003). However, since there is no highly conserved amino acid sequence and because the connector regions are highly variable, PTB domains are not easily recognized from primary sequence, despite their conserved fold.

In contrast to SH2 domains, PTB domains fail to show a highly conserved phosphotyrosine binding site.

1.2.8.2 PTB domain proteins are almost exclusively adaptor proteins

Structurally, the largest group of PTB domain proteins has an N-terminal PTB domain and is devoid of other defined protein domains (Yaffe 2002). The second largest group pairs the PTB domain with a PH domain, that is located N-terminal to the PTB domain. Other protein interaction domains are also often found in combination with PTB domain, but, with the exception of the phosphatase domain-containing tensin proteins, there are virtually no conserved catalytic domains described in PTB domain proteins. Fittingly, the PTB domain proteins are less well characterized in a functional context than SH2 proteins, although this might also result from the comparatively lower attention they received due to their late description and the smaller number of oncogenes. While the group of PH and PTB domain-containing proteins are almost exclusively thought of as adaptor proteins involved in growth factor receptor signaling, the PTB domain proteins as a whole seem to have functions ranging from cytoskeletal and focal adhesion regulation to G-protein association and activation. Interestingly, the PTB domain proteins with similar domain configuration often have diverse PTB domains, indicating cases of convergent evolution. For example, the group of proteins containing a PTB domain only comprises NUMB and DAB1 with their "Shc-like" or "Dab-like" PTB domain as well as FRS2 with its "IRS1-like" PTB domain; similarly, the group of PH and PTB domain-containing adaptors contains the eponymous IRS1 and also APPL1 and APPL2, the PTB domains of which resemble those of SHC1 and DAB1 more than that of IRS1.

1.2.9 Phosphotyrosine binding by other protein domains is rare

Aside from SH2 domains and PTB domains, several other protein domains have been shown to bind to phosphorylated tyrosine residues on other proteins (Kaneko et al. 2012b). Nevertheless, these reports seem singular and incidental.

The PH domain is closely related to the PTB domain. The most common function of PH domains is phosphoinositol binding and other common PH domain functions are often related, like phospholipid binding. Tyrosine-phosphorylated peptides are also very similar to phosphoinositol or phospholipids. Therefore, the PH domain suggests itself for non-SH2, non-PTB domain phosphotyrosine binding. Indeed, there are examples of this, although the number is low. For example, the PH domain of IQSEC1 binds EGFR-derived peptide in phosphotyrosine-dependent manner (Morishige et al. 2008) and even the name-giving PLEK PH domain specifically binds a tyrosine-phosphorylated protein (Liu & Makowske 1999). In addition to this, a number of other domains or domain arrangements have also been shown to recognize phosphorylated tyrosine residues in a protein or peptide context. For example, the C2 domains of PRKCD and PRKCQ bind a CDCP1-derived phosphotyrosine-containing peptide. Interestingly, PRKCQ, but not PRKCD, becomes activated upon binding (Stahelin et al. 2012). Canonically, C2 domains bind Ca2+ and, usually dependent on Ca2+ binding, phospholipids (Rizo & Südhof 1998). A PKM isoform binds phosphotyrosine peptides causing the release of FBP1 (Christofk et al. 2008). CBLL1 uses a novel HYB domain to bind phosphotyrosine peptides (Fujita et al. 2002) and PEBP1 binds RAF1 in a phosphotyrosine-dependent manner (Simister et al. 2011). In *Caenorhabditis elegans*, tyrosine-phosphorylated mbk-2 has been shown to be sequestered by deficient PTP domains of egg4 and egg5 (Cheng et al. 2009).

1.2.10 Protein tyrosine phosphatases are the oldest phosphotyrosine signaling components

Protein tyrosine phosphatases complete the phosphtyrosine signaling ”toolkit” (Pincus et al. 2008) by providing the final component. While there is no inherently necessary association of semantic categories and biochemical function, the tyrosine kinases are canonically considered ”writers” and the phosphatases ”erasers”. This may be mere historical accident, resulting from the much earlier description and characterization of kinases (Hunter 2009). On the other hand, the biological utilization of phosphotyrosine signaling seems to largely match this interpretation. Specifically, tyrosine kinases are usually activated in response

to specific signals, often embodied as or associated with specific signaling molecules, like growth factor proteins. Further, protein-protein interactions are usually induced (as opposed to inhibited) by tyrosine phosphorylation and often have an activating effect on downstream processes. Finally, protein tyrosine phosphatases are often inactivated following signaling events like ligand binding, either forming inhibitory intra- or intermolecular interactions or by becoming covalently but reversibly oxidized on the catalytic cysteine residue as a consequence of hydrogen peroxide release (Ostman & Böhmer 2001). Hydrogen peroxide has been recognized relatively early as an inducer of intracellular tyrosine phosphorylation (Kadota et al. 1987), but the exact and complete mechanism of action is still unclear. It seems clear, however, that, dependent on the cell type, a CYBB- or a NOX1-containing NADPH oxidase complex is one of the main producers of physiological hydrogen peroxide and that hydrogen peroxide production is a response to growth factor / receptor binding and involves PI3K action (Rhee et al. 2003). Such a mechanism of tyrosine phosphatase regulation provides ties to other processes involving hydrogen peroxide, like cytotoxic immune response or mitochondrially controlled apoptosis (Bae et al. 2011), but seems inherently unspecific. Additionally, protein tyrosine phosphatases are the evolutionarily oldest part of the phosphotyrosine signaling "toolkit", suggesting that they are the least specific components. Thus, protein tyrosine phosphatases have often been considered unspecific and less relevant for medical purposes than protein tyrosine kinases, even in spite of circumstantial evidence to the contrary, like the high number of active tyrosine phosphatases (107 in human (Alonso et al. 2004)), which rivals the number of tyrosine kinases (90 in human (Robinson et al. 2000)). However, the universality expressed in this view is currently being revised (Ramponi & Stefani 1997).

1.2.10.1 There are several groups of protein tyrosine phosphatases

Systematically, protein tyrosine phosphatases are rather diverse and usually divided into three groups: classical protein tyrosine phosphatases, dual specificity phosphatases and low molecular weight phosphatases (Zhang 1998; Andersen et al. 2001; Neel & Tonks 1997; Tonks 2006). Others (Alonso et al. 2004) group tyrosine phosphatases based on amino acid sequence of the catalytic domain and come up with four classes: class I Cys-based protein tyrosine phosphatases, in-

cluding classical protein tyrosine phosphatases and dual specificity phosphatases, class II Cys-based protein tyrosine phosphatases, including low molecular weight phosphatases, class III Cys-based protein tyrosine phosphatases and Asp-based protein tyrosine phosphatases.

1.2.10.1.1 Classical protein tyrosine phosphatases are either intramembrane receptors or cytoplasmic

Analogous to protein tyrosine kinases, classical protein tyrosine phosphatases come in two forms, transmembrane receptor phosphatases and non-transmembrane receptor phosphatases (Zhang 1998; Andersen et al. 2001; Neel & Tonks 1997; Tonks 2006). Akin to receptor tyrosine kinases, transmembrane receptor phosphatases dimerize upon ligand binding. However, in contrast to receptor tyrosine kinases, dimerization is usually accompanied by trans-binding of an "inhibitory wedge", reducing catalytic activity (Majeti et al. 2000). The ligands of transmembrane phosphatases are often extracellular matrix proteins or they form homophilic interactions, i.e. they bind the extracellular domains of identical transmembrane phosphatases expressed by neighboring cells (Ostman & Böhmer 2001). Non-transmembrane phosphatases are thought to be inactivated via receptor tyrosine kinase action, either by becoming phosphorylated and forming inhibitory intra- or intermolecular interactions, like PTPN11 (Hof et al. 1998), or as a consequence of hydrogen peroxide release.

1.2.10.1.2 Dual-specificity phosphatases are mostly specialized on MAP kinase signaling

Dual-specificity phosphatases are strongly associated with MAP kinase signaling (Keyse 2000). As indicated by the designation, dual-specificity phosphatases are capable of removing phospho groups from tyrosine, serine and threonine residues in proteins. This behavior seems sensible in the context of MAP kinase signaling, since MAP kinases auto-phosphorylate on tyrosine residues in the activation loop (Huang & Tan 2012). However, dual-specificity phosphatases are increasingly recognized as functioning outside of this context (Huang & Tan 2012). In human, there are twenty-five dual-specificity phosphatase genes, two of which (DUSP24 and DUSP27) are rendered catalytically inactive by a

cysteine-serine substitution in the active site (Huang & Tan 2012). The group is sub-classified as typical or atypical, based on the presence or absence of a CDC25 domain, respectively. This domain is generally responsible for the specific binding of MAP kinases. Interestingly, there are several dual-specificity phosphatases without this domain that have been shown to be involved in MAP kinase signaling and several dual-specificity phosphatases carrying this domain that seem to have no function in MAP kinase signaling (Huang & Tan 2012). The typical dual-specificity phosphatases are further divided into three groups based on their predominant subcellular localization. The non-MAP kinase signaling-related functions associated with dual-specificity phosphatases are mostly related to JAK/STAT signaling. Specifically, many dual-specificity phosphatases interact with and dephosphorylate STAT proteins or regulate them indirectly (Huang & Tan 2012). For many dual-specificity phosphatases, knock-out mice show phenotypes related to immune function, the most obvious field of intersection between MAP kinase signaling and JAK/STAT signaling.

1.2.10.1.3 Low molecular weight phosphatases appear to be the oldest group of protein tyrosine phosphatases

Low molecular weight phosphatases are the group most different from the other protein tyrosine phosphatases, possibly indicating a different evolutionary origin or an earlier divergence. In comparison to the other tyrosine phosphatases, they are rather small, with a typical molecular weight of about 20 kDa, are probably the most unspecific group of tyrosine phosphatases and are also prevalent in prokaryotes (Ramponi & Stefani 1997). Thus, low molecular weight phosphatases may function as unspecific antagonists of protein tyrosine kinases, counteracting ligand-independent activation of receptor tyrosine kinase signaling pathways.

1.2.10.2 The protein tyrosine phosphatases complement kinase action providing a fully adjustable signaling mechanism

In essence, efficient phosphotyrosine signaling is possible only in the interplay of tyrosine kinases, tyrosine phosphatases and phosphotyrosine-recognizing proteins. Apparently, phosphatases emerged earliest and, thus, are overall less

specific than kinases and phosphotyrosine-recognizing proteins. The tyrosine phosphatases seem to be a polyphyletic group and the evolutionarily younger members are apparently more prone to carrying out functions specific to a limited set of phosphorylation sites, while the older ones counter kinase action in a more unspecific fashion. At the very least, this mechanism guarantees that kinase activation can have constant effects over time. However, it may even be superior to more complex and more frugal (in terms of ATP consumption) systems that directly regulate phoshorylation levels when subcellular gradients of phosphorylated proteins are advantageous, provided the kinase or the phosphatase activity is properly localized. In any case, insights into phosphatase action, especially in connection with kinases, will surely prove useful for understanding how proteins interacting in a phosphotyrosine-dependent manner turn external stimuli into appropriate cellular responses.

1.3 Protein-protein interactions form networks

1.3.1 Protein interactions are more directly related to cellular function than gene sequences

The vast majority of cellular processes is regulated and carried out by proteins. Therefore, understanding what proteins do and under which conditions is vital for understanding cells on the molecular level (Cusick et al. 2005). The Human Genome Project (Venter et al. 2001) is a valuable basis for research in this direction in that it provides an inventory of possibly expressed proteins that is in principle complete. This information has to be complemented by protein expression data and protein-protein interaction information. Protein expression data is being collected in many cell types and under many different conditions in microarray and RNASeq experiments, both of which are high troughput compatible. Protein-protein interaction data are more difficult to obtain, for a number of reasons (Braun et al. 2009). First of all, protein-protein interactions are fundamentally transient and cannot be directly observed in cells that have not been tampered with in some manner. In other words, proteins have to be modified or overexpressed, or cells have to be manipulated to allow for interaction detection, e.g. by antibody binding. Additionally, native protein expression

levels differ by many orders of magnitude and protein interactions detected in a natural environment may be of a secondary nature, i.e. both proteins may bind to a third one, but not directly to each other. A way around these problems is assessing protein-protein interactions in more artificial settings, like in vitro or yeast two-hybrid assays. As a consequence, each protein-protein method detects only a relatively small subset of interactions and the results of several methods have to be combined to obtain a complete picture. At the same time, the number of possible interactions that have to be assayed is a second order function of the number of proteins of interest. Therefore, most high throughput protein-protein interaction screens operate on a gene level (ignoring isoform specificity) and post-translational modifications. Especially the latter are of extreme importance for the regulation of growth and development, however, because they are capable of signaling on a much faster time-scale and with higher efficiency.

1.3.2 Protein interaction networks are a source for biological hypotheses

Protein-protein interaction networks are a rich source for novel hypotheses regarding the biological relevance of individual proteins, the relation of biological pathways and the fundamental organization of cellular regulatory mechanisms (Barabási & Oltvai 2004). These hypotheses can be based on the network structure alone, for example, proteins with a higher number of interactions are more like to be important, as are proteins and interaction that are more central to the network (Batada et al. 2006). Even more extensive hypotheses can be reached, when the interaction network is coupled with additional information. For example, interacting proteins whose expression is highly correlated are probably members of the same complex or a protein interacting with many other proteins often found mutated in connection with a specific disease likely play a role in the development of said disease (Chuang et al. 2007; Barabási et al. 2011).

1.3.3 Low throughput follow-up experiments raise confidence in protein-protein interaction screens

Large scale protein-protein interaction screening experiments are often concluded by low throughput experiments focusing on a very small subset of interactions underlining the biological relevance of the data set (Goehler et al. 2004; Lim et al. 2006; Wong et al. 2007). At the same time, such follow-up experiments often support the quality of the data set in terms of false positive rate estimates. Unfortunately, these two purposes suggest different strategies. Investigation of biological context usually requires restriction to a very small subset of interactions, whereas demonstrating data set quality demands a large subset, if it is supposed to be nearly as meaningful as statistical bioinformatics analyses, like comparison to other datasets or interaction co-enrichment analysis. Therefore, the quality, range of suitable further applications and degree of acceptance received from the scientific community of a protein-protein interaction study profit greatly from an expedient selection of low throughput follow-up assays.

1.3.4 Biological pathways are reflected in protein-protein interactions

Protein-protein interaction networks are strongly related to biological pathway maps (Scott et al. 2006). Nevertheless, there are substantial differences. First and foremost, protein-protein interaction maps focus on biophysical information and lead to biological function as an emerging property of a complex system in a bottom-up fashion. Biological pathway maps, on the other hand, focus on a specific biological function or subsystem and aim to explain the internal workings of the system in question in terms of interacting components in a top-down approach. As a consequence, protein-protein interaction networks may contain interactions whose function is unknown, or, in the extreme case, that have no biological function, and biological pathway maps may contain links that have no direct biophysical correlate. Nevertheless, in an ideal world (with perfect interaction networks and pathway maps), the pathway map should be traceable in terms of physical interactions, because transfer of information between proteins without physical interaction is extremely unlikely.

Thus, it almost seems like a necessary consequence that the view of biological regulation is shifting more and more from individual, isolated pathways to strongly interconnected pathways and, finally, regulatory networks. This development is especially prominent when it comes to pathways involved in the regulation of growth and development, because there is a striking overlap in the components involved, leading to the question: how can growth factor receptor pathways achieve specific outcomes employing similar sets of proteins (Chao 1992)?
Unbiased, genome-wide screening results are especially valuable due to the high potential for finding functional connections otherwise concealed (Stelzl & Wanker 2006). Protein-protein interaction networks therefore grant additional clarity above all, where biological semantics deviate from one-to-one relationships between genes or proteins and functions. Some examples of such situations are (a) where direct experimental analysis is impossible because the gene in question has other functions that interfere with the assay, for example, because mutants are not viable, (b) where conventional approaches report inconclusive results due to cooperative effects, for example as a result of partial compensation by another gene with overlapping function, or (c) where emergent properties are concerned, i.e. when the complex system develops properties that are different from the combined properties of its parts.

1.3.5 Protein-protein interaction networks improve classification of cancers

A very direct link from fundamental research to medical application is the use of protein interaction networks for the diagnosis, comprehension and treatment of disease (Barabási et al. 2011). A prime example of a complex disease that is difficult to understand because of its plasticity, is cancer (Hong & Hait 2010). In fact, it has been argued, that cancer should be regarded as a multitude of diseases, and the number of different types of cancer is usually estimated as more than one hundred (Hanahan & Weinberg 2000). Generally, cancers derived from the same type of tissue are often similar. Therefore, a classification scheme based on tissue type is widely used. On the other hand, cancers derived from the same tissue are also often very different, resulting in a more specific

classification scheme, for example, lung cancer is divided into small cell and non-small cell lung cancer and breast cancer is commonly classified based on the expression levels of ESR1/2, PGR and ERBB2 or the mutational state of BRCA1/2. Other times, it is even argued that every cancer is different (Greaves & Maley 2012). Therefore, analysis of cancer development is a prime candidate for the application of network-based methods.
A prime example for the application of protein interaction network-based methods for cancer research is the work of Chuang et al. (2007), who overlayed a protein-protein interaction network with information about expression changes in breast cancer and extracted discriminative subnetworks. They further demonstrated that these subnetworks are more accurate and more robust classifiers of metastasis than those based on single genes. Pujana et al. (2007) combined PPI and expression data to identify novel breast cancer genes and prove the validity of their method by way of example on the HMMR gene. Taylor et al. (2009) analysed coexpression differences between hubs and their interaction partners in breast cancer to find diagnostic network modules. Nibbe et al. (2009) combined PPI and expression data to identify subnetworks discriminative for colon cancer. Li et al. (2012) used nearest neighbor-, GO- and shortest path-based methods to derive novel colon cancer genes from known colon cancer genes and protein-protein interaction information. Interesting about all of these examples is the restriction to relatively well-known and homogenous forms of cancer, like breast and colon cancer, and the motivation for the selection of the additional data. Almost all of the authors explicitly state that they aim to attach a dynamic element to the protein-protein interaction maps. This indicates a demand for protein interaction maps that have intrinsic dynamic elements, especially for the analysis of more complex types of cancer.

1.3.6 Assays of dynamic interactions are focused on small subsets of protein-protein combinations

Protein-protein interactions are generally detected in artificial settings, either in vitro, in artificial cellular environments, or as tagged and overexpressed versions (Braun et al. 2009). Additionally, protein-protein detection methods have high false negative rates. As a consequence, protein-protein interaction maps gener-

ally contain all detected interactions without context information (Rinner et al. 2007). In principle, protein-protein interactions that can be detected by affinity purification/mass spectrometry-based methods are amenable to dynamic analysis, because they allow comparing results from different cells, states etc. In practice, there are several studies exemplifying the possibilities, but have a very limited scope in terms of interacting proteins, especially bait genes. For example, Blagoev et al. (2003) used SILAC to compare proteins precipitated from HeLa cells with the GRB2 SH2 domain before and after EGF stimulation; Brand et al. (2004) used an ICAT-LC/MS approach to compare co-immunoprecipitation partners of Mafk in murine erythroleukemia cells before and after DMSO-induced differentiation; Blagoev et al. (2004) used SILAC to compare proteins precipitated from HeLa cells with a phosphotyrosine-specific antibody (4G10) before and after 1 min and 10 min of EGF stimulation; Wu et al. (2006) compared EGFR binders in A 431 cells before and after 0.5 min, 2 min, 10 min and 240 min of EGF stimulation; Wang & Huang (2008) used (mixing after purification)-SILAC to differentiate stable and dynamic RPN11 binders, Mousson et al. (2008) compared MAP (mixing after purification)-SILAC and PAM-SILAC results to differentiate stable and dynamic TBP binders in HeLa cells and Kito et al. (2008) used a similar method for eIF2Bβ(GCD7 locus) binders in *S. cerevisiae*. A very noteworthy example is the study presented by Bisson et al. (2011). Here, affinity precipitation(AP)-coupled selected reaction monitoring (SRM) was used to identify dynamic binders of GRB2 in HEK293T cells. Binders were classified as SH2- and SH3-dependent with the help of domain mutations. Additionally, EGF, FGF, HGF, IGF, INS and PDGFB stimulation were compared, and for EGF stimulation, samples from five different time points were assessed. Nevertheless, not one of these studies employs more than one bait gene. Thus, a comprehensive study addressing the dynamics of interactions among full-length proteins experimentally, is currently missing.

1.4 Protein-protein interactions detected by different methods are complementary

The detection of biologically relevant protein-protein interactions is difficult due to the high number and diversity of proteins and their interactions. In contrast

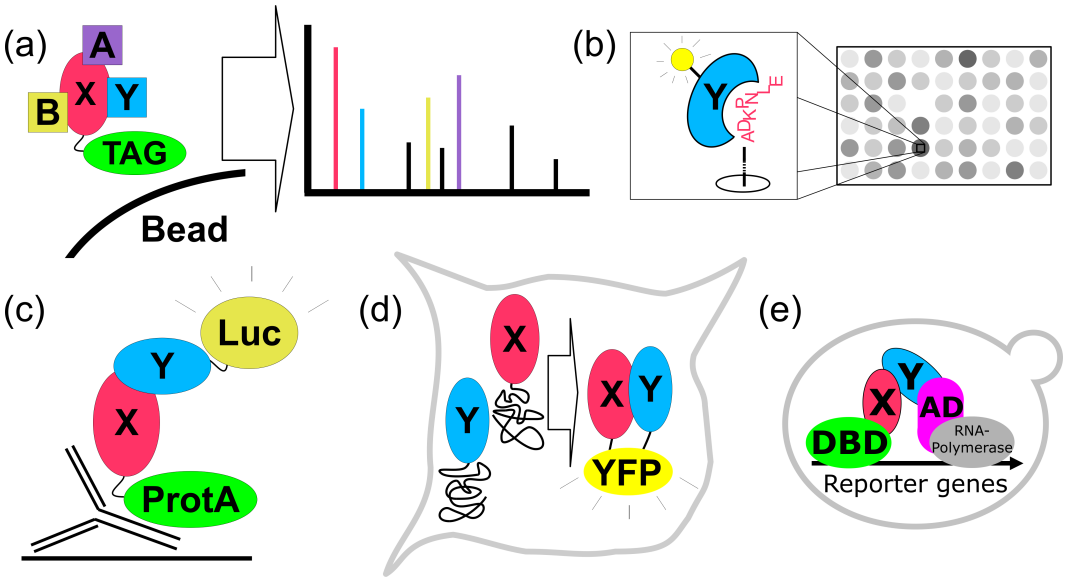


Figure 7: Complementary methods for detecting protein-protein interactions use different molecular mechanisms. Schematic depicting the most common methods for detecting protein-protein interaction. In affinity purification-coupled mass spectrometry-based methods (a) complexes of proteins containing a usually tagged protein of interested are purified and member proteins are identified in mass spectrometric analyses. In peptide array binding assays (b) spots of peptides are immobilized on a membrane and probed with the protein of interest; after washing, bound proteins are visualized either directly or indirectly. In high throughput co-immunoprecipitation (c) two tagged proteins of interested are expressed in mammalian culture and immobilized in a microplate using the tag fused to one of the proteins; binding is determined by measuring the activity of the luciferase tag fused to the other protein. In protein complementation assays (d) a suitable reporter protein is split and fused to two proteins of interest; if the two proteins interact, the reporter is reconstituted when the two fusion proteins are expressed in mammalian cell culture. In yeast two-hybrid experiments (e) two proteins of interest are fused to the DNA-binding and the RNA-polymerase activation doamin, respectively, and expressed in a yeast cell containing one or more reporter gene constructs; interaction of the two proteins activates the reporter genes, usually conferring the ability to grow on selection medium.

to other macromolecules carrying biological functions, like RNA and DNA, proteins are made up from a high number of building blocks with versatile biochemical properties. As a consequence, protein-protein interaction surfaces and interaction parameters, like k_{on} and k_{off} rates or dissociation constants, are highly diverse, making protein-protein interactions almost impossible to predict from protein structure alone and each protein-protein detection method

suitable for a limited subset of interactions only. Thus, protein-protein knowledge relies heavily on "Professoral intuition"-driven low throughput experiments and repeated unbiased high throughput experiments with different methods, the results of which are combined in a complementary fashion. Accordingly, there are many methods for the detection of interactions between proteins (Figure 7).

1.4.1 Affinity purification-coupled mass spectrometry is the main method for detecting stable protein complex associations

A powerful method for the detection of protein-protein interactions, especially in the context of protein complexes is affinity purification coupled with mass spectrometry (Aebersold & Mann 2003; Bauer & Kuster 2003). Typically, a bait protein construct containing an affinity purification tag is overexpressed in the cells of interest and this tag is utilized in purifying the bait protein and its binders. The binders are then prepared for mass spectrometric analysis and identified (Figure 7a).

Commonly used purification tags are TAP, FLAG, Protein A, HA or Myc tag, but other purification methods, like His-Nickel-NTA or glutathione-GST purification, or antibodies against endogenous proteins can also be used. well-established tag systems are preferred for high throughput experiments, because they provide experimental procedures amenable to parallelization and equalize expression level and purification efficiency (Lichty et al. 2005).

Almost all mass spectrometric analysis methods rely on peptide cleavage by trypsin, followed by a database comparison (Fenyö 2000). Specifically, sequences obtained by exact mass determination of b- and y-ions, specific combinations of precursor and product ions (SRM, Lange et al. 2008) or characteristic co-occurrence of tryptic peptides (MASCOT, Perkins et al. 1999) are used to select proteins from a database prepared for the respective analysis.

Due to the enormous advances in mass spectrometry technology in the last decades, mass spectrometry-based method are very effective tools for detecting protein-protein interactions. Nevertheless, these methods are far from perfect or universally applicable (Aebersold & Goodlett 2001). First of all, because complexes are analyzed, the detected interactions are not binary, i.e. it is unclear

whether the interactions are direct or mediated by other members of the complex. Secondly, due to the huge differences in protein expression rates, interactors in low abundance can easily be masked by highly abundant interactors or even non-specific binders. Additionally, many weak or transient binders are lost in the washing steps and proteins that do not produce suitable peptides when treated with trypsin cannot be detected at all.

1.4.2 Peptide array binding is an efficient method for detecting protein-peptide interactions

Another method that can be used to discover protein-protein interaction is peptide array binding (Volkmer et al. 2012). For this type of experiment, many different peptides are spotted on a surface, typically a glass slide, incubated with a purified tagged protein, and subsequently with a fluorescently labeled antibody (Figure 7b). Even though this method requires purified proteins, it is very powerful, because a very high number of peptides can be assayed at the same time. The biggest drawback of this method is the use of peptides instead of full-length proteins. Many protein-protein interactions rely on complex surfaces, not just linear peptide motifs (Chakrabarti & Janin 2002).

1.4.3 Luciferase makes co-immunoprecipitation amenable to high throughput approaches

The LUMIER method is a high throughput adaptation of an established method for the detection of protein-protein interactions, co-immunoprecipitation (Barrios-Rodiles et al. 2005). Classical co-immunoprecipitation experiments employ two antibodies. The first one is used to precipitate the first protein of interest from a cell lysate. The second one demonstrates the binding of the second protein of interests on a Western blot. In LUMIER type experiments, the first protein is bound to the surface of a 96-well microplate and the second protein is fused to a luciferase gene. Luciferases are enzymes that can activate specific substrates to produce bioluminescence (Figure 7c). This way, the co-immunoprecipitation can be performed in 96-well format and read out in a luminescence reader, making LUMIER-type methods suitable for high throughput protein-protein interaction screens.

1.4.4 Protein complementation assays visualize protein interactions in intact mammalian cells

Protein complementation assays utilize proteins that can be easily observed and can be split into two parts that regain their function when brought together (Kerppola 2006). The two parts are fused to two proteins of interest and the resulting hybrid proteins are expressed in cell culture. If the two proteins interact, the reporter protein is reconstituted and becomes active, for example, a split-GPF becomes fluorescent or a split-luciferase produces luminescence (Figure 7d). A big advantage of this method is that the interaction can be observed in intact cells and in a specific subcellular localization. On the other hand, the method requires a relatively large manual effort to ward against false-positives resulting from signals produced by misfolded proteins finding each other in the endoplasmic reticulum.

1.4.5 The yeast two-hybrid systems is the prime method for detecting binary protein-protein interactions in screening assays

The yeast two-hybrid system (Fields & Song 1989) is a very well-established method for the detection of protein-protein interactions. The proteins that are assayed for interaction are fused to transcription factor domains, the so-called bait proteins to a DNA-binding domain and the so-called prey proteins to a transcriptional activation domain. These hybrid proteins are overexpressed in a yeast cell containing one or several appropriate reporter gene constructs. If the two proteins interact, the reporter gene activation provides an easily observable signal (Figure 7e).

Typical reporter genes are genes encoding enzymes involved in the synthesis of fundamental cellular anabolites, like amino acids or nucleotides, employed in respective auxotroph mutant yeast strains. In this case, the yeast can grow on medium lacking these anabolites only if the two hybrid proteins interact. Another commonly used reporter gene is β-galactosidase. Observing this signal requires an additional reporter gene assay in which the yeast cells are opened and treated with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (XGAL) to

produce a blue stain where the proteins interact. The vast majority of yeast two-hybrid systems employs GAL4 or LexA DNA-binding and transcriptional activation domains.
Historically, the majority of yeast two-hybrid experiments relied on complex prey libraries. These libraries were often cDNA libraries generated by reverse transcription using oligo(dT) primers. As a consequence, the results of the early yeast two-hybrid experiments are biased towards interactions mediated by C-terminal proteins domains and, more importantly, towards highly expressed genes. Therefore, more thorough investigations employ prey matrices instead, i.e. arrays of defined prey cultures. This setup requires a higher initial effort, but has many advantages. The most import ones are equal testing probabilities for all prey strains and defined testing parameters allowing repeating and comparing of individual assays.
Yeast two-hybrid provides binary interactions at an extremely low false-positive rate (Braun et al. 2009; Venkatesan et al. 2009), making it the method of choice for screening applications without distinctive prior estimation. Thus, we applied the yeast two-hybrid system as the primary screening method in this study.

1.4.5.1 The yeast two-hybrid system is inherently well-suited for high throughput applications

In addition to its inherent ability to produce high quality interactions in terms of low false positive rate, the yeast two-hybrid system has several other virtues making it extremely suitable for screening assays. The most important ones are the ones making it amenable to high throughput studies (Parrish et al. 2006). Specifically, the robustness of experimental procedures especially related to the screening itself allows working in microplate format and using robots to combine and process bait and prey strains, thereby combating the combinatorial explosion effect.

1.4.5.2 Analysis of phosphotyrosine-dependent protein interactions requires introduction of a tyrosine kinase gene

The analysis of protein phosphotyrosine-related effects in *S. cerevisiae* is intrinsically artificial, since there is no phosphotyrosine signaling in yeast. As a consequence, there is virtually no background in terms of protein tyrosine

kinase and protein tyrosine phosphatase activity (Gnad et al. 2009). Therefore, every important component related to tyrosine phosphorylation, like the kinase, has to be provided by the experimental setup. This implies that native regulatory mechanisms common to organisms that make use of phosphotyrosine signaling are absent in the yeast environment. In turn, this makes the experiments cleaner and the attribution of modifying and recognizing proteins easier in several respects. In mammalian cells, in contrast, protein tyrosine kinase activity can never be abolished completely, because that would cause the cells early demise. Furthermore, even if it were possible, experimental validation of the absence of kinase activity would be practically impossible. In addition, kinases often activate other kinases, which makes actually determining which kinase phosphorylates a specific site or enables a specific interaction extremely difficult.

Modified yeast two-hybrid systems have been used to detect phosphotyrosine-dependent protein-protein interactions several times before (Keegan & Cooper 1996; Dombrosky-Ferlan & Corey 1997; Osborne et al. 1995; Rocchi et al. 1998; Marti et al. 1998; Delahaye et al. 2000; Warner et al. 2000; Yamada et al. 2001; Cao et al. 2002; Sayós et al. 2004; Ingley et al. 2006; Sylvester et al. 2010), but, to our knowledge, there has never been a comprehensive screen for phosphotyrosine-dependent interactions based on the yeast two-hybrid technology.

1.5 Aim of the study

Proteins are the most important functional units on the molecular level of biological systems. As a consequence, understanding protein-protein interactions is key to understanding how cells work. Phosphotyrosine-dependent protein-protein interactions are especially interesting, because they are extremely important in regulating growth and development and can be highly dynamic. The close connection to regulation of growth and development makes phosphotyrosine signaling extremely relevant for many fields of interest in current biomedical research, like cancer or immunological disease.

This study attempts to fill a gap in protein-protein interaction knowledge by presenting an unbiased genome-scale set of phosphotyrosine-dependent protein-protein interactions found in yeast two-hybrid experiments. In doing so, we seek to gain insight into the architecture of the central cellular regulatory network in terms of general, structural make-up as well as identifying specific proteins fulfilling important network functions.

The current knowledge of phosphotyrosine-dependent protein-protein interactions stems mainly from low throughput experiments based on serendipitous observations or high throughput studies relying on peptide microarray or complex purification/mass spectrometry techniques.

We address this issue by screening a comprehensive set of genes with protein domains strongly implicated in protein phosphotyrosine recognition against a genome-scale prey matrix using an extended version of a well-established yeast two-hybrid system and methodology. We mean to assess the quality of our results in independent coimmunoprecipitation experiments from mammalian cell culture. Complementary to this experimental validation we are going to use computational analyses to test the data set for consistency and estimate its novelty and completeness.

Following the successful validation of our data, we will demonstrate its power as basis for hypothesis generation by exploring a small number of hypotheses prompted by our screen that are particularly interesting in that they improve upon current paradigms of phosphotyrosine signaling.

To sum up, we aim to assemble a comprehensive set of phosphotyrosine-dependent protein-protein interactions that is largely complementary to the current

knowledge, because we employ a yeast two-hybrid approach. Further, we mean to demonstrate the data sets high quality using a number of established validation tools.

Chapter 2

Material and Methods

2.1 Buffers and reagents

2.1.1 Vendors

All materials used for this study were acquired from well-established vendors through the purchase department of the Max-Planck-Institute for Molecular Genetics. The vendors are detailed here. In the remainder of this work, vendors are referred to by the short names.

Vendor (Short Name)	**Details**
Abcam	Abcam plc 330 Cambridge Science Park Cambridge CB4 0FL, UK `http://www.abcam.com`
Agilent	Agilent Technologies Sales & Services GmbH & Co.KG Life Sciences & Chemical Analysis Hewlett-Packard-Str. 8 D-76337 Waldbronn `http://www.agilent.com/chem`
Applichem	AppliChem GmbH Ottoweg 4 D-64291 Darmstadt `http://www.applichem.com`

Vendor (Short Name)	Details
BD Biosciences	Becton, Dickinson and Company 1 Becton Drive Franklin Lakes, New Jersey 07417-1880 `http://www.bd.com`
BioTez	BioTeZ Berlin Buch GmbH Robert-Rössle-Strasse 10 Haus D 79 3.OG D-13125 Berlin `http://www.biotez.de`
Bio-Rad	Bio-Rad Laboratories Headquarters 1000 Alfred Nobel Drive Hercules, CA 94547 `http://www.bio-rad.com`
Eppendorf	Eppendorf Vertrieb Deutschland GmbH Peter-Henlein-Str. 2 50389 Wesseling-Berzdorf `http://www.eppendorf.com`

Fermentas	Fermentas UAB
	Fermentas
	LT-02241 Vilnius
	V. Graiciuno 8
	`http://www.fermentas.com`
	now fully owned subsidiary of:
	Thermo Fisher Scientific Inc.
	81 Wyman Street Waltham, MA USA 02451
	`http://www.thermofisher.com`

GE Healthcare	GE Healthcare Europe GmbH Munzinger Str. 5 D-79111 Freiburg `http://www.gelifesciences.com`
Greiner Bio-One	Greiner Bio-One North America, Inc., 4238 Capital Drive Monroe, NC 28110 USA `http://us.gbo.com`

Vendor (Short Name)	Details
Invitrogen	brand of: Life Technologies GmbH Frankfurter Straße 129B D-64293 Darmstadt www.lifetechnologies.com now fully owned subsidiary of: Thermo Fisher Scientific Inc. 81 Wyman Street Waltham, MA USA 02451 http://www.thermofisher.com
Jackson ImmunoResearch	Jackson ImmunoResearch Europe Ltd. Unit 7, Acorn Business Centre, Oaks Drive Newmarket, Suffolk, CB8 7SY, UK http://jireurope.com/
Merck	Merck KGaA Frankfurter Straße 250 D-64293 Darmstadt http://www.merck.de
Millipore	Millipore Corporation 290 Concord Road, Billerica Massachusetts, USA http://www.millipore.com now fully owned subsidiary of: Merck KGaA Frankfurter Straße 250 D-64293 Darmstadt http://www.merck.de
MWG	MWG-Biotech AG Anzinger Str. 7a D-85560 Ebersberg http://www.mwg-biotech.com/
NEB	New England Biolabs 240 County Road Ipswich, MA 01938-2723 USA https://www.neb.com/

Vendor (Short Name)	Details
Perkin Elmer	PerkinElmer LAS (Germany) GmbH Ferdinand-Porsche-Ring 17 D-63110 Rodgau `http://www.perkinelmer.de/`
Promega	Promega Corporation 2800 Woods Hollow Road Madison, WI 53711 USA `http://www.promega.com`
Roche	Roche Diagnostics GmbH Sandhofer Straße 116 D-68305 Mannheim `http://www.roche.de`
Roth	CARL ROTH GMBH + CO. KG Schoemperlenstr. 1-5 D-76185 Karlsruhe `http://www.carlroth.de`
Santa Cruz	Santa Cruz Biotechnology, Inc. 10410 Finnell Street Dallas, Texas 75220 USA `http://www.scbt.com`
Serva	SERVA Electrophoresis GmbH Carl-Benz-Str. 7 D-69115 Heidelberg `http://www.serva.de`
Sigma	Sigma-Aldrich Chemie GmbH Riedstraße 2 D-89555 Steinheim `http://www.sigmaaldrich.com`
Thermo Fisher	Thermo Fisher Scientific Inc. 81 Wyman Street Waltham, MA USA 02451 `http://www.thermofisher.com`
TPP	TPP Techno Plastic Products AG Zollstrasse 7 CH-8219 Trasadingen `http://www.tpp.ch`

2.1.2 Consumables

Deionized water (H_2O) was produced in house using a Barnstead Nanopure Diamond Life Science (Thermo Fisher) water system or a multi-stage distillation system assembled by the technical department, and had a specific resistance of at least 18 MΩ/cm. Additionally, for all microbiological applications, it was autoclaved.

The gene-specific oligodeoxyribonucleotides used for generating new Entry clones were purchased from MWG. All other oligodeoxyribonucleotides used in this study were obtained from BioTez. All oligodeoxyribonucleotides were order at 10 nmol scale and RP-HPLC purified.

Simple, standardized chemicals, i.e. those commonly referred to by IUPAC nomenclature, were ordered in *per analysis* quality from either Applichem, Merck, Promega, Roth, Serva or Sigma, according to the current best offer, and were considered, for all practical purposes, pure and identical.

The other consumables used in this study are listed here:

Plastic labware:

Name (according to vendor)	**Vendor**	**ArticleID (according to vendor)**
384 Well Microplate, PS, F-Bottom	Greiner Bio-One	781186
96 Well Microplate, PS, F-Bottom	Greiner Bio-One	655101
Deepwell Plates 96 / 2000 µl	Eppendorf	0030.501.3XX
Lid for Microplates, PS	Greiner Bio-One	656190
Nunc Low Profile BioAssay Dish	Thermo Fisher	240845
OmniTray Single Well w/Lid Sterile PS	Thermo Fisher	242811
Tissue Culture Test Plate 24-Well F	TPP	92024
96 Well Flat Top PCR Platte	Biozym	710875

For cloning:

Name (according to vendor)	**Vendor**	**ArticleID (according to vendor)**
Gateway BP Clonase II Enzyme mix	Invitrogen	11789100
Gateway LR Clonase II Enzyme mix	Invitrogen	11791100
Deoxynucleotide (dNTP) Solution Set	NEB	N0446S
Phusion Hot Start High-Fidelity DNA Polymerase	NEB	F-540L
QuikChange Site-Directed Mutagenesis Kit	Agilent	200518

For microbiology and related assays:

Name (according to vendor)	Vendor	ArticleID (according to vendor)
Adenine hemisulfate salt	Sigma	A9126
Agar	BD Biosciences	214010
Betaine	Sigma	B2629
L-Histidine	Sigma	H8125
L-Leucine	Sigma	L8912
L-Tryptophan	Sigma	T0254
Peptone	BD Biosciences	211677
Tryptone	BD Biosciences	211705
Uracil	Sigma	U0750
Yeast Extract	BD Biosciences	212750
Yeast Nitrogen Base without Amino Acids	BD Biosciences	291920
Polyethylene Glycol 6000	Applichem	A1387
Salmon Sperm DNA	Sigma	D1626
Hybond-N+ Chargen Nylon Membrane	GE Healthcare	RPN2250B
X-Gal	Fermentas	R0402

For Western Blotting and related experiments:

Name (according to vendor)	Vendor	ArticleID (according to vendor)
Anti-Phosphotyrosine Antibody clone 4G10	Millipore	05-321X
Bovine Serum Albumin, Fraction V (BSA)	Sigma	A7906
Coomassie Brilliant Blue R 250	Serva	17525
Glass beads	Sigma	G4649
MagicMark Western Protein Standard (20-120 kDa)	Invitrogen	LC5600
Nitrocellulose	Bio-Rad	162-0115
Rotiphorese Gel 40 (37,5:1)	Roth	T802.1
UltraPure TEMED	Invitrogen	15524010
Western Lightning Plus-ECL	Perkin Elmer	NEL104001EA

For mammalian cell culture:

Name (according to vendor)	Vendor	ArticleID (according to vendor)
Dulbecco's Modified Eagle Medium	Invitrogen	31966047
Dulbecco's phosphate-buffered saline (PBS)	Invitrogen	14190094
Lipofectamine 2000 Transfection Reagent	Invitrogen	11688019
Opti-MEM	Invitrogen	31985062
Triton-X 100	Serva	39795

For Co-Immunoprecipitation assays:

Name (according to vendor)	Vendor	ArticleID (according to vendor)
AffiniPure Rabbit Anti-Sheep IgG (H+L)	Jackson ImmunoResearch	313-005-003
Bright-Glo	Promega	E2620
cOmplete, Mini, EDTA-free	Roche	11836170001
Phosphatase Inhibitor Cocktail 1	Sigma	P2850
Phosphatase Inhibitor Cocktail 2	Sigma	P5726
Doxycycline hyclate	Sigma	D9891

For protein complementation assays:

Name (according to vendor)	Vendor	ArticleID (according to vendor)
Poly-D-lysine hydrobromide	Sigma	P1149
DAPI	Roth	6335
Depex	Serva	18243
NEO-Clear	Merck	1098435000
Rhodamine Phalloidin	Invitrogen	R415

For immunofluorescent analysis:

Name (according to vendor)	Vendor	ArticleID (according to vendor)
Anti-TSPAN2 antibody	Abcam	ab77105
Bovine Serum Albumin, Fraction V	Sigma	A7906
DyLight 488-conjugated AffiniPure Donkey Anti-Rabbit IgG (H+L)	Jackson ImmunoResearch	711-485-152
DyLight 488-conjugated AffiniPure Rabbit Anti-Mouse IgG (H+L)	Jackson ImmunoResearch	315-485-003
GRB2 Antibody (C-23)	Santa Cruz	SC-255
Purified Mouse Anti-Human Retinoblastoma Protein Antibody	BD Biosciences	554136

2.1.3 Solutions and media for yeast growth experiments

Glucose and supplements for auxotrophic selection were prepared and autoclaved separately:

- 20 x Glucose stock solution: 1 mol glucose was dissolved in 500 ml H_2O and autoclaved.

- 100 x histidine stock solution: 0.8 g L-histidine were dissolved in 400 ml H_2O and autoclaved.
- 100 x adenine stock solution: 0.8 g adenine hemisulfate salt were dissolved in 400 ml H_2O and autoclaved.
- 100 x uracil stock solution: 0.8 g uracil were dissolved in 400 ml H_2O and autoclaved.
- 100 x leucine stock solution: 4 g L-leucine were dissolved in 400 ml H_2O and autoclaved.
- 100 x tryptophan stock solution: 0.8 g L-tryptophan were dissolved in 400 ml H_2O and autoclaved.
- Copper(II) sulphate was prepared as 200 mM stock solution and sterilized by filtration.

Liquid medium was prepared as stock solution without glucose or supplements for auxotrophic selection:

- 1.25 x YPD stock solution: 5 g yeast extract and 10 g peptone were dissolved in 400 ml H_2O and autoclaved.
- 1.25 x NB stock solution: 3.35 g Yeast Nitrogen Base without Amino Acids (BD Biosciences) were dissolved in 400 ml H_2O and autoclaved.
- 1.25 x NBG stock solution: 3.35 g Yeast Nitrogen Base without Amino Acids (BD Biosciences) and 29.44 g betain were dissolved in 250 ml 62.5% glycerol and autoclaved.

Before use, 25 ml 20 x glucose stock solution and 5 ml 100 x stock solution of each desired supplement were added to the desired medium stock solution, then they were filled up to 500 ml with H_2O and mixed. NBG medium was used for freezing and storing at -80°C. Yeast cultures were frozen up to two times in this medium.
Solid medium was prepared either from stock solutions, similar to liquid medium, or directly, in a ProfiClav PC20B (biomedis Vertriebsgesellschaft mbH, D-35394 Giessen). For stock solutions agar and Yeast Nitrogen Base without Amino Acids (BD Biosciences) were autoclaved separately:

- 1.25 x YPD agar stock: 5 g yeast extract, 10 g peptone and 10 g agar were dissolved in 400 ml H_2O and autoclaved.
- 2.5 x NB stock solution: 6.7 g Yeast Nitrogen Base without Amino Acids (BD Biosciences) were dissolved in 400 ml H_2O and autoclaved.
- 2.5 x Agar: In a 500 ml bottle, 10 g agar were dissolved in 200 ml H_2O and autoclaved.

Before use, 25 ml 20 x glucose stock solution and 5 ml 100 x stock solution of each desired supplement were added to each stock solution, then they were filled up to 500 ml with H_2O. For medium with 20 µM Cu^{2+} 50 µl 200 mM $CuSO_4$ stock solution were added. Afterwards, the agar was solubilized by heating to 90°C in a microwave oven and mixed by swirling carefully. The agar solution was allowed to cool down to 60°C before casting. For preparation of YPD solid medium with the ProfiClav PC20B, for 1 l of medium, 10 g yeast extract, 20 g peptone, 20 g agar and 950 ml H_2O were added, autoclaved and cooled to 60°C. Once 60°C were reached, 50 ml 20 x glucose stock solution were added, allowed to be mixed thoroughly and cast. For preparation of NB solid medium with the ProfiClav PC20B, for 1 l of medium, 13.4 g Yeast Nitrogen Base without Amino Acids (BD Biosciences) 40 g agar and H_2O were added, autoclaved and cooled to 60°C. The amount of water was chosen to amount to 1 l after addition of glucose and supplements. Once a temperature of 60°C was reached, 50 ml 20 x glucose stock solution and 10 ml 100 x stock solution of each desired supplement were added, as well as 100 µl 200 mM $CuSO_4$ for Cu^{2+} medium, allowed to be mixed thoroughly and cast. NB and NBG media were referred to according to the supplements added and left out. Each of the five supplements was represented by a single letter:

H	histidine
A	adenine
U	uracil
L	leucine
T	tryptophane

Supplements that were left out were listed first, preceded by a minus sign, followed by a slash, then the supplements added. The order of supplements was

H, A, U, L, T on both sides of the slash. Addition of 20 µM Cu^{2+} was indicated as "(Cu^{2+})".

2.2 Plasmids

The plasmids used in this study are listed here with short descriptions and relevant characteristics:

Name	**Reference / Source**	**Function**	**Markers**
pDONR221	Invitrogen	Gateway Entry	Kanamycin (*Escherichia coli*)
pDONR223	(Rual et al. 2004)	Gateway Entry	Spectinomycin (*E. coli*)
pBTM116-D9	(Worseck et al. 2012)	yeast two-hybrid bait expression	Tetracycline (*E. coli*) Tryptophan (*S. cerevisiae*)
pACT-4-DM pGAD426-D3	(Worseck et al. 2012)	yeast two-hybrid prey expression	Ampicillin (*E. coli*) Leucine (*S. cerevisiae*)
pASZ-C-DMtet	(Stotz & Linder 1990)	yeast two-hybrid kinase expression	Tetracycline (*E. coli*) Adenine (*S. cerevisiae*)
pASZ-CN-DM	(Stotz & Linder 1990)	yeast two-hybrid kinase expression with NLS	Ampicillin (*E. coli*) Adenine (*S. cerevisiae*)
pPAReni-DM	(Palidwor et al. 2009)	Protein A-fusion expression in mammalian cell culture	Tetracycline (*E. coli*)
pFireV5-DM	(Palidwor et al. 2009)	luciferase-fusion expression in mammalian cell culture	Ampicillin (*E. coli*)
pVEN-F1N-DM pVEN-F1C-DM pVEN-F2N-DM pVEN-F2C-DM	created in our laboratory, based on YFP PCA fusion expression vectors in (Remy et al. 2004)	protein fragment complementation assay construct expression	Ampicillin (*E. coli*)

As Gateway plasmids, all of these contain a chloramphenicol resistance gene and a ccdB gene, as long as they contain the "death cassette". The ccdB gene

will kill any *E. coli* cells with wildtype gyrase unless the "death cassette" has been replaced by an ORF.
The yeast two-hybrid prey matrix also contained classical, i.e. non-Gateway, plasmids. For a description of the prey matrix see (Suter et al. 2013).

2.3 Cells

2.3.1 Yeast strains

In this study, three well-established haploid *S. cerevisiae* strains were used as basis for mating and subsequent yeast two-hybrid experiments. All three are incapable of switching mating types and auxotrophic for adenine, tryptophan and leucine, which is exploited for selecting strains transformed with pASZ, pBTM, or pGAD/pACT plasmids, respectively. The resulting diploid strains contain a lexA-based reporter gene system for growth on medium without histidine and uracil and expression of β-galactosidase. Genotypes are provided in (Goehler et al. 2004). An overview of the strains and the types of plasmids they were transformed with for the assays presented in this study is provided here:

Strain	**Mating Type**	**transformed with (for)**
L40c	MATa	pBTM & pASZ (screening) pASZ (kinase plate assay)
L40ccU2	MATa	pBTM & pASZ (screening) pASZ (kinase plate assay)
L40ccα	MATα	pGAD/pACT (screening) pBTM & pGAD/pACT (kinase plate assay)

2.3.2 Mammalian cell lines

Mammalian cell culture experiments presented here were based exclusively on derivatives of HEK293 cells. These cells were initially created by transforming primary human embryonic kidney cells with adenovirus DNA (Graham et al. 1977). Specifically, "T-Rex" cells were obtained from Invitrogen. Where induction of a kinase was desired, a version with a stably integrated doxycycline-inducible

ABL2 ORF was employed. This version was created in our laboratory using the "Flip-In" system provided by Invitrogen.

2.4 Gateway cloning

High throughput molecular biological experiments often require testing the same ORFs in several different plasmids. For example, in this study, all bait ORFs were also used as prey. In addition, examining interactions in mammalian cell culture required sub-cloning the respective ORFs into plasmids allowing the expression of Protein A, luciferase or split-YFP fusion proteins in mammalian cells. To facilitate the sub-cloning and make it amenable to high throughput approaches, we employed the Gateway system (Invitrogen). This system relies on the lambda phages enzymes for genomic integration and excision and the respective attachment sites. Except for a small number of prey plasmids, all plasmids used in this study were Gateway plasmids. Each ORF was either obtained as a Gateway Entry clone or such a clones was produced by PCR and Gateway BP cloning. The plasmids used for functional experiments were subsequently generated by Gateway LR cloning. When an ORF was desired in two Gateway Destination plasmids with different *E. coli* resistance markers, up to two Gateway LR reactions were combined.

2.4.1 Generation of Gateway Entry clones

For ORFs not available as Gateway Entry clones, PCR primers specific to the ORF were generated and used to amplify the ORF sequence while introducing invariable flanking sequences:

1 µl	2 ng/µl	template DNA (plasmid, see below)
2.5 µl	10 µM	forward and reverse primer mix (each)
1 µl	10 mM	dNTP mix (each)
10 µl	5 x	Phusion buffer
1.5 µl		DMSO
0.5 µl	2 U/µl	Phusion polymerase
33.5 µl		H_2O

The PCRs were processed in a thermocycler according to this protocol:

Initial denaturation:
98°C 30 seconds
Elongation (25 cycles):
98°C 8 seconds
55°C 23 seconds
72°C 2 minutes
Final elongation:
72°C 8 minutes
Cool down:
4°C until removed from machine

In a second PCR step, these flanking regions were extended to form functional attB sites:

1	µl		template DNA (PCR products, diluted 1:100 with H_2O)
2.5	µl	10 µM	attB1for & attB2rev (each)
1	µl	10 mM	dNTP mix (each)
10	µl	5 x	Phusion buffer
1.5	µl		DMSO
0.5	µl	2 U/µl	Phusion polymerase
33.5	µl		H_2O

The PCRs were processed in a thermocycler using the same protocol as for the first PCR.
The PCR products were purified by PEG/$CaCl_2$ precipitation, as recommended by the manufacturer:

1. For each ORF, 25 µl PCR were added to 75 µl TE (10 mM Tris-HCl pH 8, 1 mM EDTA) in a fresh tube and mixed by vortexing.
2. 50 µl 30% PEG 8000 / 30 mM $MgCl_2$ Solution (Invitrogen) were added to each tube and mixed by vortexing.
3. The PCR products were centrifuged for 15 min at 25°C and 15,000 xg.
4. The supernatant was removed with a pipette and discarded.

5. The pellets were dissolved in 10 µl TE, each.

The purified PCR products were used as substrate in a Gateway BP reaction:

3 µl	$MgCl_2$/PEG purified PCR product	
1 µl	75 ng/µl	pDONR221
1 µl	5 x	Gateway BP Clonase II Enzyme Mix (Invitrogen)

Each reaction was incubated at 25°C for 18 h and directly used for *E. coli* transformation. Each Gateway Entry clone generated this way was verified by Sanger sequencing.
Non-Gateway-compatible plasmids were ordered from Imagenes (now Invitrogen) and Gateway Entry clones were generated for these genes, using these templates and primers:

GeneID	Gene	IMAGE clone	Plasmid	GenBankID
321	APBA2	IRALp962P1059Q	pOTB7	BC082986
	forward primer	AAAAAGCAGGCTTAATGGCCCACCGGAAGC		
	reverse primer	AGAAAGCTGGGTCCTACTAGATGTACAGCGGGG		
9479	MAPK8IP1	IRATp970G0988D	pBluescriptR	BC068470
	forward primer	AAAAAGCAGGCTTAATGGCGGAGCGAGAAAG		
	reverse primer	AGAAAGCTGGGTCCTACTACTCCAGGTAGATATC		
5355	PLCG1	IRAKp961M0798Q	pCMV-SPORT6	BQ876810
	forward primer	AAAAAGCAGGCTTAATGGCGGGCGCCGC		
	reverse primer	AGAAAGCTGGGTCCTACTAGAGGCGGTTGTCTC		
79890	RIN3[1]	IRAKp961I19168Q	pBluescriptR	BC070062
	forward primer	AAAAAGCAGGCTTAATGATCCGACACGCCG		
	reverse primer	AGAAAGCTGGGTCCTATCACAGGAAGTTGGGCTC		
55620	STAP2	IRAKp961M034Q	pCMV-SPORT6	BC000795
	forward primer	AAAAAGCAGGCTTAATGGCCTCTGCCCTGAG		
	reverse primer	AGAAAGCTGGGTCCTATCAGTGCTCCAGTGCC		
7410	VAV2	DKFZp686B22207Q	pSPORT1_Sfi	BX640754
	forward primer	AAAAAGCAGGCTTAGACGGGGTCCTTCTGTG		
	reverse primer	AGAAAGCTGGGTCCTATCACTGGATGCCCTCC		

[1]Note that for RIN3 two different Gateway Entry clones were obtained this way, both of which were successfully mapped to the RIN3 gene.

2.4.2 Generation of assay-specific plasmids by Gateway LR recombination

To employ in different assays, ORFs available as Gateway Entry clones could be moved to appropriate Gateway Destination vectors in simple and high throughput amenable Gateway LR recombination reactions:

1 µl	200 ng/µl	Gateway Entry clone DNA
3 µl	25 ng/µl	Gateway Destination plasmid(s) (total concentration)
1 µl	5 x	Gateway LR Clonase II Enzyme Mix (Invitrogen)

Each reaction was incubated at 25°C for 18 h and directly used for *E. coli* transformation. Where two Gateway Destination vectors were used, *E. coli* strains were grown to completion in LB liquid medium and tested for transformation with multiple plasmids by placing a drop (>1 µl) of this culture on an agar plate containing the other respective antibiotic. Strains growing in this test were discarded.
All of the Gateway LR recombination reactions with the Destination plasmids pPAReni-DM or pFireV5-DM were performed by Nouhad Benlasfer.

2.5 Yeast Two-Hybrid

2.5.1 Yeast transformation

The first step in most yeast two-hybrid experiments is the creation of yeast strains containing the yeast two-hybrid plasmids with the ORFs of interest. A high throughput-adapted version of the lithium acetate method (Gietz & Woods 2002) was used this study:

1. A 3 ml tube of YPD liquid medium was inoculated with the yeast strain that was supposed to be transformed from a freshly grown agar plate and incubated at 30°C until the next day. In case of strains already containing a kinase plasmid (pASZ-C or pASZ-CN), -A/HULT medium was used instead of YPD.

2. On the next day, 2 ml of the over-night culture were transferred to a 1 l Erlenmeyer flask without baffles containing 30 ml YPD liquid medium preheated to 30°C and incubated at 30°C shaking vigorously. In case of strains already containing a kinase plasmid (pASZ-C or pASZ-CN), -A/HULT medium was used instead of YPD.

3. After 3 h 30 min, the yeast culture was transferred to a 50 ml tube and centrifuged for 5 min at 2000 rpm in an Eppendorf A-4-81 centrifuge rotor.

4. The supernatant was discarded and the pellet was resuspended in 10 ml TE (10 mM Tris-HCl pH 8, 1 mM EDTA).

5. The cells were centrifuged again for 5 min at 2000 rpm an Eppendorf A-4-81 centrifuge rotor.

6. The supernatant was discarded and the pellet was resuspended in 1.1 ml 100 mM $LiCH_3COO$, 0.5 x TE, 1 M Sorbitol and incubated for 15 min at 25°C.

7. In an individual well of a PCR plate for each transformation reaction, 500 ng plasmid DNA were mixed with 15 µg Salmon Sperm DNA (Sigma) in 4 µl 0−1 x TE. In case of transformation with two plasmids at once (for the kinase plate assay), the amount of plasmid DNA was increased to 1 µg, 500 ng per plasmid.

8. 11 µl yeast cell suspension were added to each well and mixed by vortexing carefully.

9. 58 µl 100 mM $LiCH_3COO$, 1 x TE, 40% PEG 6000 were added to each well and mixed by vortexing carefully.

10. The PCR plate was incubated at 30°C for 30 min.

11. 8 µl DMSO were added to each well and mixed by vortexing carefully.

12. The PCR plate was incubated at 42°C for 7 min.

13. For the desired number of copies, a drop (∼5 µl) of yeast solution from each well was placed on an agar plate selective for the plasmid or plasmids

that were transformed with, using a pipette or pin tool, allowed to dry and incubated at 30°C for 3 days.

14. Before use for functional assays, the generated yeast strains were transferred to NBG medium and replicated on agar plates.

2.5.2 Generation of Yeast Two-Hybrid Bait strains

Yeast two-hybrid bait strains were generated by transforming haploid MATa strains with pASZ plasmids first, and, in a second step, transforming these strains with pBTM plasmids. Transformations of pASZ plasmid-containing strains with pBTM plasmids selected for the second round of screened were performed by Petra Birth.

1. The haploid MATa yeast strains L40c and L40ccU2 were transformed with the nine non-receptor tyrosine kinase ORFs selected for screening, each in both, pASZ-C and pASZ-CN plasmid backbones, as well as empty pASZ-C and pASZ-CN plasmids.

2. The resulting yeast strains were transformed with the pBTM-D9 plasmids of the ORFs that were selected as bait for screening. For the second round, only the strains carrying the three kinase ORFs selected for screening in the second round and the empty pASZ-C and pASZ-CN plasmids were transformed.

3. The bait strains were re-arrayed to produce 96-well plates in which pBTM116-D9 inserts were identical in each well in the same row and pASZ version (pASZ-C or pASZ-CN) and inserts were identical in each well in the same column. Further, all strains containing identical inserts in the pBTM116-D9 plasmid were arranged in the same row and all strains containing identical pASZ version (pASZ-C or pASZ-CN) and inserts were arranged in the same column in all plates, with the exception of the empty pASZ-C and pASZ-CN plasmids, which were present in three columns, each. The eight pBTM116-D9 inserts in each plate were supposed to form the pools for screening and retesting (assuming the absence of growth in the autoactivation test). Thus, for the second round of screening all ORFs for a GeneID were placed in the same 96-well plates.

2.5.3 Autoactivation test

Before any yeast two-hybrid screen, each individual bait strain was tested for autoactivation. Each bait strain was mated with a prey strain containing an empty pACT4 plasmid, i.e. a plasmid from which only the GAL4 activation domain is expressed. Any bait strains growing on selective agar lacking histidine and uracil were disqualified from further yeast two-hybrid experiments, i.e. they were not used for screening and disregarded in the evaluation of retest experiments. The autoactivation tests of bait strains prepared for the second round of screening were performed by Petra Birth.
This is the protocol that was followed for the autoactivation test:

1. The bait strains were stamped on -AT/HUL NB agar plates in 96-well format and incubated for 3 days at 30°C.

2. Two days after stamping the bait strains, a -L/HAUT NB liquid culture of prey strain containing an empty pACT4 plasmid was inoculated and incubated at 30°C until the next day shaking vigorously in an Erlenmeyer flask without baffles with a nominal volume of at least five times the culture volume. The volume of the culture was 120 ml plus 12 ml per 96-well plate of bait strains.

3. The liquid culture was distributed to 96-well flat-bottom plates (120 µl per well).

4. The bait strains were transferred from the -AT/HUL NB agar plates to the microplates, mixed and stamped on YPD agar using a pin tool.

5. The YPD agar plates were incubated at 30°C until the next day.

6. Yeast spots were transferred from YPD agar plates to 96-well flat-bottom plates containing 120 µl -ALT/HU NB medium in each well, mixed and stamped on -ALT/HU NB agar plates using a pin tool.

7. The -ALT/HU NB agar plates were incubated at 30°C for 2 days and controlled for complete growth.

8. The yeast spots were transferred from the -ALT/HU NB agar plates to 96-well flat-bottom plates containing 120 µl -ALT/HU NB medium in

each well, mixed and stamped on -HAULT(Cu^{2+}) NB agar plates using a pin tool.

9. The -HAULT(Cu^{2+}) NB agar plates were incubated at 30°C for 5 days and evaluated. Each bait strain corresponding to a spot growing on -HAULT(Cu^{2+}) NB agar was disqualified from further yeast two-hybrid experiments, i.e. not used for screening and disregarded evaluating retest experiments.

2.5.4 Yeast Two-Hybrid Screening

For genome-scale screening, bait strains were tested in pools comprising all combinations of up to eight ORFs representing phosphotyrosine-recognizing domain-containing genes in pBTM116-D9 and up to nine non-receptor tyrosine kinase ORFs in pASZ-C or pASZ-CN (one pBTM116-D9 and one pASZ plasmid per strain; see Appendix XXX for exact combinations screened). Each bait pool was mated individually with each prey strain available in our laboratorys prey matrix at the time (note that a substantial number of prey strains became available between the first and the second round of screening; see Appendix XXX for exact combinations screened) up to four times using independently transformed bait strains. Combinations of bait pools and prey strains growing were considered primary hits potentially representing protein-protein interactions. Most of the screening using L40c-derived bait strains in the first round and all of the screening in the second round were performed by Petra Birth.
In detail, yeast two-hybrid screening experiments followed this protocol:

1. For each bait pool, the bait strains were stamped on -AT/HUL NB agar plates and incubated for 3 days at 30°C.

2. On the following day, for each bait pool, one copy of the prey matrix was stamped on -L/HAUT NB agar plates and incubated for 3 days at 30°C.

3. The fully grown bait strains were transferred from the -AT/HUL NB agar plates to 96-well flat-bottom plates containing 120 µl -AT/HUL NB liquid medium per well using a pin tool and mixed by vortexing carefully.

4. Eight 96-well deepwell plates containing 1.2 ml -AT/HUL NB liquid medium per well were inoculated for each bait pool by adding 5 µl medium

from one well of the 96-well flat-bottom plates containing the bait strains. The first microplate for each bait pool, five deepwell plates were inoculated. The other three deepwell plates were inoculated with the left half (columns 1-6) of the second microplate for the bait pool, twice. For this purpose, the left half of the deepwell plate was inoculated with the left half of the 96-well flat-bottom plate, the deepwell plate was turned by 180° and inoculated again. Positions corresponding to autoactive bait strains were not inoculated in this step by removing the respective tips before transferring the 5 µl yeast suspension to the deepwell plates. (The right half of the second 96-well plates, containing the pASZ plasmid without kinase ORFs, were not included in the primary screen.)

5. The deepwell plates were incubated at 30°C shaking vigorously until the next day.

6. For each bait pool, all eight deepwell plates were vortexed until any yeast accumulated at the bottom of the wells was completely suspended, brought together in a 5 l beaker and mixed by swirling.

7. The mixture of bait strains was distributed to 384-well flat-bottom plates, adding 43 µl to each well. The number of microplates filled corresponded to the number of 384-well plates comprising the prey matrix, plus one (for control purposes).

8. The prey strains were transferred from the -L/HAUT NB agar plates to the 384-well plates containing the bait strains, mixed and stamped on YPD agar plates using a pin tool.

9. The YPD agar plates were incubated at 30°C until the day after the next.

10. The yeast spots were transferred from the YPD agar plates to 384-well flat-bottom plates containing 43 µl -ALT/HU NB medium in each well, mixed and stamped on -HAULT(Cu^{2+}) NB agar plates using a pin tool.

11. The -HAULT(Cu^{2+}) agar plates were incubated at 30°C for 5 days and evaluated. For each spot, the corresponding prey strain and bait pool were noted as a primary hit.

2.5.5 Retesting of primary Yeast Two-Hybrid hits

Potentially interacting protein pairs producing yeast growth in the screening experiments were only considered interacting if they could be reproduced in an independent retest experiment. In addition to making the interaction data more reliable and stringent, the retest experiment provides specific information regarding which bait strains produced the signal identifying the interaction partners and allowing a distinction between phosphotyrosine-dependent and independent interactions. It also reveals autoactive prey strains. To achieve a high positive retest rate, only combinations of bait pools and non-autoactive prey strains with at least two primary hits were included in the retest experiments. Autoactive prey strains were identified using growth statistics collected in the laboratory of Erich Wanker at the MDC Berlin, as well as ours. All of the retesting of primary hits following screening with L4oc-derived bait strains in the first round and most of the retesting following the second round of screening were performed by Petra Birth.
In detail, yeast two-hybrid screening experiments followed this protocol:

1. Each combination of bait pool and prey strain with at least two primary hits was included in the retest set.

2. Each combination involving an autoactive prey strain was removed from the retest set. For this purpose, a prey strain was considered autoactive if it grew in more than 9 of 24 primary screens related to prior studies performed in our laboratory and either fulfilled the retesting criteria in more than 17 of 19 compound screens performed in the laboratory of Erich Wanker at the MDC Berlin or there was no information about it available from this source.

3. For each bait pool, the 96-well plates were combined into 384-well plates by transferring two independent transformations of the first 96-well plate into the first and fourth quadrant and two transformations of the second 96-wells plate into the second and third quadrant.

4. For each combination of bait pool and prey strain in the retest set, one copy of the corresponding 384-well bait plate was stamped on -AT/HUL NB agar and incubated at 30°C for 3 days.

5. After two days, 30 ml -L/HAUT NB liquid medium per combination in the retest list were inoculated with the respective prey strain and incubated at 30°C shaking vigorously until the next day.

6. On the next day, the -L/HAUT NB liquid prey cultures were distributed to 384-well flat-bottom plates, adding 43 µl to each well, one microplate per 30 ml liquid medium.

7. For each combination in the retest list, the bait strains from one copy of the respective 384-well bait plate were transferred from the -AT/HUL NB agar plates to the 384-well flat-bottom plates containing the respective prey strain, mixed and stamped on -ALT/HU NB agar plates using a pin tool.

8. The -ALT/HU NB agar plates were incubated at 30°C for 3 days and controlled for complete growth.

9. Yeast spots were transferred from -ALT/HU NB agar plates to 384-well flat-bottom plates containing 43 µl -ALT/HU NB medium in each well, mixed and stamped on -HAULT(Cu^{2+}) NB agar plates using a pin tool.

10. The -HAULT(Cu^{2+}) NB agar plates were incubated at 30°C for 5 days and evaluated. Each spot growing on -HAULT(Cu^{2+}) NB agar represented one combination of interacting proteins and kinase. Each combination of bait and prey plasmids corresponded to two rows, identical kinase plasmids to two diagonally adjacent positions with the two rows. As a consequence, the following pattern were used to derive protein-protein interactions on an Entrez GeneID level (ignoring autoactive bait strains):

Pattern	**Interpretation**
Growth on the whole plate	autoactive prey strain (no interaction)
Fully grown pair of rows	independent interaction
Pair of rows grown in several kinase-containing positions only	phosphotyrosine-dependent interaction
single spots	late autoactivity (no interaction)

2.5.6 Assessment of Tyrosine Phosphorylation parameters

To assess the influence of copper sulfate to the interaction-selective -HAULT agar plate, diploid yeast two-hybrid strains can be transferred to -HAULT agar with and without added copper sulfate.

1. A yeast two-hybrid experiment was performed as described for the retest experiments up to the last stamping step.

2. Instead of the last stamping step, the yeast spots were transferred from the -ALT/HU NB agar plates to 96-well flat-bottom plates containing 120 µl -ALT/HU NB medium in each well, mixed by vortexing carefully and spotted on -ALT/HU, -HAULT and -HAULT(Cu^{2+}) NB agar by pipetting 5 µl yeast suspension onto the surface of each. For β-galactosidase activity read-out, the yeast suspension was spotted on -HAULT and -HAULT(Cu^{2+}) NB agar with a piece of Nylon membrane (GE Healthcare Hybond-N+), additionally.

3. The agar plates were incubated for 3 days and evaluated, either directly or by β-galactosidase activity determination. (Note that the yeast spots can be used for Western blotting.)

2.5.7 Kinase interaction specificity analysis

2.5.7.1 Generation of catalytically inactive kinase ORFs

Protein-protein interactions allowing growth in yeast two-hybrid experiments in the presence of an active tyrosine kinase, but not in its absence, could be explained by other mechanisms than phosphotyrosine binding. For example, the kinase might bind to the bait and prey proteins simulating the effects of direct binding, a concept often referred to as ”bridging”. To be able to refute these hypotheses, catalytically inactive kinase versions were generated by introducing a mutation to a highly conserved lysine residue in the ATP-binding cassette, which, when changed to a methionine residue, renders the kinase inactive. The mutations were introduced by PCR using the QuikChange Site-

Directed Mutagenesis Kit (Agilent) for all kinases that were used for screening or retesting:

ORF	RefSeq RNA	ORF		RefSeq	
Primer	Sequence	Start	End	Start	End
ABL2	NM_005158.4	1	3504	46	3549
forward	cagccttacagttgctgtgATGacattgaaggaagatacc			930	969
reverse	ggtatcttccttcaatgtCATcacagcaactgtaaggctg			969	930
BMX (ORF #1)	NM_001721.6	1	2026	112	2137
forward	gtatgatgttgctgttATGatgatcaaggagggctcc			1428	1464
reverse	gtctctcgacaagaCATcaccgccaccagg			1464	1428
FES	NM_002005.3	1	2469	97	2565
forward	cctggtggcggtgATGtcttgtcgagagac			1851	1880
reverse	ggtatcttccttcaatgtCATcacagcaactgtaaggctg			1880	1851
FRK	NM_002031.2	1	1518	448	1965
forward	ccactccagtagcagtgATGacattaaaaccaggttcaatg			1214	1254
reverse	cattgaacctggttttaatgtCATcactgctactggagtgg			1254	1214
FYN (ORF #1)	NM_002037.5	1	1613	608	2220
forward	gaaacacaaaagtagccataATGactcttaaaccaggcac			1482	1521
reverse	gtgcctggtttaagagtCATtatggctacttttgtgtttc			1521	1482
JAK2	NM_004972.3	1	3397	495	3891
forward	gggaggtggtcgctgtaATGaagcttcagcatagtac			3121	3157
reverse	gtactatgctgaagcttCATtacagcgaccacctccc			3157	3121
PTK2 (ORF #1)	NM_005607.4	59	2043	1474	3458
forward	cagctttggcggttgcaattATGacatgtaaaaactgtacttcg			1639	1682
reverse	cgaagtacagtttttacatgtCATaattgcaaccgccaaagctg			1682	1639
SYK (ORF #1)	NM_003177.6	1	1908	256	2163
forward	gtgaaaaccgtggctgtgATGatactgaaaaacgaggcc			1441	1479
reverse	ggcctcgtttttcagtatCATcacagccacggttttcac			1479	1441
TNK1	NM_003985.4	1	1986	233	2218
forward	ccagtggctgtcATGtccctccgggtag			662	689
reverse	ctacccggagggaCATgacagccactgg			689	662

In this protocol, complimentary PCR primers carrying the mutated version of the site that is supposed to be changed and matching the nucleotide residues surrounding the site are used for a PCR replicating the complete template plasmid. Afterward, the template is removed by incubating with DpnI, a restriction endonuclease specifically recognizing dam methylated DNA. Since the

mutation is introduced in the Gateway Entry vectors, the common problem of additional mutations resulting from the comparatively low fidelity of DNA replication by PCR (compared to within intact *E. coli* cells) is minimized. Each reaction was set up like this:

19.5 µl		H_2O
2.5 µl	10 x	QuikChange reaction buffer
0.5 µl	50 x	QuikChange dNTP mix
1 µl	5 ng/µl	template DNA
0.5 µl	10 µM	forward primer
0.5 µl	10 µM	reverse primer
0.5 µl	50 x	PfuTurbo Polymerase

and was processed in a thermocycler according to this protocol:

Initial denaturation:		
95°C	30 seconds	
Elongation (25 cycles):		
95°C	30 seconds	
55°C	60 seconds	
68°C	5 / 6 minutes	(BMX, FRK, FYN, PTK2, SYK, TNK1 / ABL2, JAK2, FES)
Final cooling:		
4°C	until removed from machine	

Upon completion, 0.5 µl DpnI (Agilent) were added to each PCR sample, carefully mixed in several slow aspirate-dispense cycles with a pipette. Subsequently, the samples were incubated at 37°C for 1 hour and stored at -20°C for bacterial transformation. All generated mutations were verified by Sanger sequencing. All catalytically deficient kinase mutant were generated by Sylvia Wowro instructed by Arndt Großmann as part of her undergrad education.

2.5.7.2 Kinase Plate Assay

For a comprehensive analysis of kinase interaction specificity in the yeast two-hybrid system, haploid yeast strains containing pASZ-C and pASZ-CN plasmids with ORFs for 31 non-receptor tyrosine kinase genes, including wildtype and

catalytically inactive versions of the ORFs used for screening, were arranged in 96-well format with five empty pASZ-C and five empty pASZ-CN plasmids and mated with yeast strains containing yeast two-hybrid bait and prey plasmids known to allow yeast growth in a kinase-dependent manner. Each interaction was tested twice with independently transformed kinase plates, one based on L40c, the other on L40ccU2, by placing 5 µl drops of diploid yeast suspension on -HAULT(Cu^{2+}) agar.
The kinase plate assay was performed according to this protocol:

1. The haploid MATα yeast strain L40ccα was transformed with combinations of yeast two-hybrid bait and prey plasmids, i.e. a pBTM116-D9 and either a pGAD or a pACT plasmid, known to allow yeast growth in a kinase-dependent manner.

2. The haploid MATa yeast strains L40c and L40ccU2 were transformed with 43 non-receptor tyrosine kinase ORFs selected for screening, each in pASZ-C and pASZ-CN, as well as empty pASZ-C and pASZ-CN plasmids, five times each. The non-receptor tyrosine kinase ORFs were selected to cover 31 non-receptor tyrosine kinase genes and include catalytically inactive versions of the nine ORFs used for screening. (Note that for JAK1, the only non-receptor tyrosine kinase gene not covered by the ORF selection, several plasmids have been obtained and sequenced; none of these could be confirmed; we believe that all of them derive from the same incorrectly annotated MGC clone.)

3. The resulting kinase strains were arrayed in 96-well format in two plates identical in all aspects except parent strain. Similarly, within each plate, the left and the right half, i.e. columns 1-6 and 7-12, respectively, were identical in all aspects except plasmid backbone, i.e. pASZ-C or pASZ-CN. In addition, the ORFs were arranged to avoid placing two highly active kinases next to each other, as far as information was available. Where such information was unavailable, the ORFs carrying mutations in the ATP-binding cassette and the empty plasmids were assumed to be inactive and for the remaining kinases members of the same family were assumed to display correlated levels of activity.

4. The kinase yeast strains were stamped on -A/HULT NB agar plates and incubated at 30°C for 3 days. Each 96-well plate was stamped once for each MATα yeast strain to be tested.

5. After two days, 60 ml -LT/HAU NB liquid medium each were inoculated with the MATα yeast strains in an Erlenmeyer flasks without baffles and incubated at 30°C until the next day shaking vigorously.

6. The liquid cultures were distributed to two 96-well flat-bottom plates, each (120 µl per well).

7. The kinase strains were transferred from the -A/HULT NB agar plates to the microplates, mixed and stamped on YPD agar using a pin tool.

8. The YPD agar plates were incubated at 30°C until the next day.

9. The yeast spots were transferred from the YPD agar plates to 96-well flat-bottom plates containing 120 µl -ALT/HU NB medium in each well, mixed and stamped on -ALT/HU NB agar plates using a pin tool.

10. The -ALT/HU NB agar plates were incubated at 30°C for 2 days.

11. The -ALT/HU NB agar plates were controlled for complete growth.

12. The yeast spots were transferred from the -ALT/HU NB agar plates to 96-well flat-bottom plates containing 120 µl -ALT/HU NB medium in each well.

13. The microplates were mixed by vortexing carefully and spotted on selective -HAULT(Cu^{2+}) NB agar plates by pipetting 5 µl liquid medium onto the surface of an agar plate for each well.

14. The -HAULT(Cu^{2+}) NB agar plates were incubated at 30°C for 5 days and evaluated.

2.5.8 β-Galactosidase Assay

The yeast two-hybrid system employed in this study involves three reporter genes and offers two kinds of protein interaction read-out. The histidine and uracil reporters can be queried simply by testing for growth on selective medium. The

third reporter gene, β-galactosidase, expresses an enzyme capable of hydrolyzing β-galactosides into monosaccharides. It also accepts the analog 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) as a substrate producing galactose and 5-bromo-4-chloro-3-hydroxyindole. Under oxidative conditions, the latter forms 5,5′-dibromo-4,4′-dichloro-indigo, which is blue, making the reaction immediately observable.
In detail, the following protocol was used for the β-galactosidase assay:

1. Nylon membranes (GE Healthcare Hybond-N+) were placed on interaction-selective agar plates (-HAULT(Cu^{2+})).

2. After spotting on interaction-selective agar plates as described for the retest experiments, diploid yeast two-hybrid cultures were spotted on the nylon membranes and incubated at 30°C for several days until colonies were fully grown, but not any longer.

3. Membranes were carefully removed from the agar plates and transferred to liquid nitrogen.

4. Thoroughly frozen membranes were removed from the nitrogen and allowed to reach at least 15°C.

5. Completely thawed membranes were again frozen by transferring to -80°C and stored until the next day or up to a week.

6. Membranes were allowed to thaw at 25°C.

7. For each piece of membrane, two sheets of 3MM chromatography paper (Whatman), slightly larger than the piece of membrane were stacked in an empty agar dish and soaked with β-galactosidase activity detection solution (750 µM X-Gal, 10 mM DTT, 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$). X-Gal and DTT were added immediately before every experiment (from 20 g/l dimethylformamide and 1 M aqueous stock solutions, respectively).

8. Membranes were places onto the X-Gal-soaked papers and incubated at 37°C for several hours until blue staining was developed. Since the color-producing reaction is oxidative, the time at which the staining

develops varies with DTT freshness. Therefore, the exact incubation time was determined by monitoring the reaction.

9. When the staining was developed, the reaction was stopped by removing the membranes from the agar dishes and allowing to dry by evaporation.

2.6 Western Blot analysis of tyrosine-phosphorylated proteins

To visualize kinase activity in the yeast environment, native yeast proteins tyrosine-phosphorylated by the overexpressed human tyrosine kinase were separated by SDS-PAGE (12%), immobilized on a nitrocellulose membrane and recognized by the well-characterized pan-protein phosphotyrosine 4G10 antibody.
In detail, the following protocol was employed for the Western blot:
Preparation of yeast protein extracts:

1. A yeast two-hybrid experiment was performed as described for the retest experiments up to the last stamping step.

2. Instead of the last stamping step, the yeast spots were transferred from the -ALT/HU NB agar plates to 96-well flat-bottom plates containing 120 µl -ALT/HU NB medium in each well, mixed by vortexing carefully and spotted on -ALT/HU, -HAULT and -HAULT(Cu^{2+}) NB agar by pipetting 5 µl yeast suspension onto the surface of each.

3. The agar plates were incubated for 3 days.

4. For each yeast spot, a 1.5 ml tube containing 50 µl SDS loading buffer (2% SDS, 90 mM Tris/HCl pH 6.8, 10% glycerol, 0.2% 2-mercaptoethanol) and a spatula tip full of glass beads was prepared.

5. The yeast spots were carefully transferred to the 1.5 ml tubes using an inoculation loop.

6. Each tube was heated up to 95°C, vortexed three times for one second on the highest setting, incubated for 2 min at 95°C, vortexed three times for one second on the highest setting, incubated for another 3 min at 95°C,

vortexed three times for one second on the highest setting, and allowed to coll down to 25°C.

Preparation of polyacrylamide gels for electrophoresis:

7. Two mini-gel scaffolds (Bio-rad Mini-PROTEAN 1 mm) were assembled for casting.

8. In a beaker, 10 ml separating gel solution were prepared by adding

4.4	ml		H_2O
100	µl	10%	SDS solution
2.5	ml	1.5 M	Tris/HCl pH 8.8
3	ml	40%	acrylamide/bisacrylamide solution (37.5:1)

9. The solution was mixed by swiveling carefully.

10. The polymerization reaction was started by adding 50 µl 10% APS solution and 5 µl TEMED.

11. The solution was mixed by swiveling carefully and immediately slowly poured into the mini-gel scaffolds, leaving room for the combs plus 5 mm for the stacking gel.

12. Each gel was overlaid with 300 µl 1-butanol equalized with separating gel buffer. (1-butanol was equalized by adding an equal amount of 375 mM Tris/HCl pH 8.8, 0.1% SDS, mixing by shaking and allowing the phases to separate.)

13. The gels were allowed to polymerize by leaving them undisturbed for 1 hour.

14. The 1-butanol was removed by decanting.

15. In a beaker, 5 ml stacking gel solution were prepared by adding

3.7	ml		H_2O
50	µl	10%	SDS solution
625	µl	1 M	Tris/HCl pH 6.8
3	ml	40%	acrylamide/bisacrylamide solution (37.5:1)

16. The solution was mixed by swiveling carefully.

17. The polymerization reaction was started by adding 25 µl 10% APS solution and 2.5 µl TEMED.

18. The solution was mixed by swiveling carefully and immediately slowly poured into the mini-gel scaffolds, enough to fill the space between the glass plates completely after addition of the combs.

19. One ten-well comb was inserted carefully into each scaffold and polymerization was allowed for by leaving the gels undisturbed for 1 hour.

Separation of protein according to size by discontinuous SDS PAGE:

20. Both gels were set up in a Mini-PROTEAN Tetra Cell (Bio-rad) by placing them in the same gel holder cassette, adding running buffer (25 mM Tris, 1.44% glycine, 0.1% SDS), carefully removing the combs and rinsing the wells (by quickly pipetting running buffer from the reservoir into the wells).

21. Equal amounts of each sample were added to both gels in identical order. For yeast protein extract samples, 10 µl were used for each lane; for protein size standard (Invitrogen MagicMark), 5µl were used for each lane.

22. Proteins were electrophoretically separated by applying a constant current of 70 mA (35 mA per gel) until the buffer fronts reached the bottom of the gels.

23. Electrophoresis was stopped, the gels were retrieved, the stacking gels were removed and discarded and the separating gels were used for Western blotting and Coomassie staining.

Unspecific protein staining with Coomassie:

24. The first gel was transferred to a tray with Coomassie staining solution (50% methanol, 10% acetic acid, 0.25% Coomassie Brilliant Blue G-250) and incubated for 1 hour at 25°C slowly shaking in an undulating fashion.

25. The staining solution was removed by decanting, rinsing with destaining solution (30% methanol, 10% acetic acid), then by incubating in destaining solution until destained at 25°C slowly shaking in an undulating fashion. The destaining solution was refreshed whenever thoroughly blue.

Detection of protein tyrosine phosphorylation by Western blotting:

26. The second gel was transferred to blot buffer (48 mM Tris, 39 mM glycine, 1.3 mM SDS, 20% methanol) and incubated for 2 min.

27. One piece of nitrocellulose membrane, slightly larger than the gel, and two pieces of 3 mm chromatography paper (Whatman), both of equal size and slightly larger than the piece of membrane, were soaked in blot buffer.

28. The two pieces of chromatography paper, the membrane and the gel were stacked in a Trans-Blot SD Semi-Dry Transfer Cell (Bio-rad) in the order chromatography paper, membrane, chromatography paper, from bottom to top. The creation of air bubbles was carefully avoided, as was any unnecessary touching of the membrane.

29. The proteins were transferred to the membrane by applying a constant voltage of 20 V for 25 min.

30. The membrane was transferred to 2% BSA in TBS (25 mM Tris, 150 mM NaCl, 0.555% HCl) and incubated until the next day at 4°C slowly shaking in an undulating fashion.

31. The BSA solution was removed by decanting and replaced by 10 ml TBS containing 5 µl 4G10 antibody solution (1 g/l, Millipore).

32. The membrane was incubated with the antibody solution for 90 min at 4°C slowly shaking in an undulating fashion.

33. The membrane was washed three times with 15 ml TBS-T (1% Tween 20 in TBS) for 5 min each.

34. The membrane was washed three times with 15 ml TBS for 5 min each.

35. The membrane was incubated in 10 ml TBS containing 4 µl HRP-conjugated sheep anti-mouse antibody solution (0.55 g/l, GE Healthcare) for 60 min at 4°C slowly shaking in an undulating fashion.

36. The membrane was washed three times with 15 ml TBS-T (1% Tween 20 in TBS) for 5 min each.

37. The membrane was washed three times with 15 ml TBS for 5 min each.

38. For documentation, HRP activity was triggered using Western Lightning reagent (PerkinElmer) according to the manufacturers instructions.

2.7 Protein co-immunoprecipitation from mammalian cell culture

A subset of the protein-protein interactions found in yeast two-hybrid experiments was assayed by high throughput co-immunoprecipitation. Towards this end, the respective constructs were subcloned into the pPAReni-DM or pFireV5-DM plasmids by Gateway LR cloning, allowing expression of Protein A-tagged or firefly luciferase-tagged proteins in mammalian cell culture. For each combination of constructs analysed, three wells of a 96-well microplate with ABL2-overexpressing T-Rex (HEK293) cells were transfected with pPAReni and pFireV5 plasmids containing the respective ORFs, induced with doxycyclin or control medium on the next day and finally evaluated after another day. To do this, the cells were washed, lysed and the Protein A-tagged protein was precipitated in an IgG-coated microplate. The amount of co-immunoprecipitated luciferase-tagged protein was determined by measuring the luciferase activity using the BrightGlo kit (Promega). Input activity was determined by measuring luciferase activity in the lysate before precipitation. Background signal was measured using cells not transfected with luciferase constructs. Input activity was used to control assay quality. Average backround signal was subtracted from each measurement. Corrected measurements were averaged and compared to the respective control, i.e. non-binding protein pairs involving the same firefly luciferase-tagged construct, or, for mutation analyses, the respective wildtype construct, based on a simple multiplicative model. Differences of a factor of

at least two and with a Z-score of at least two were considered significant and substantial. The experimental part of the co-immunoprecipitation assay was performed by Nouhad Benlasfer. Detailed protocol:

1. Inoculation was begun by plating 2.5 x 10^4 ABL2 T-Rex cells in six wells in 75 µl DMEM of a 96-well cell culture plate for each combination of Protein A-tagged and firefly luciferase-tagged constructs, and incubating at 37°C in a 5% CO_2 atmosphere until the next day.

2. For each combination of constructs, a transfection master mix theoretically sufficient for seven wells was prepared from two pre-mixes:

DNA mix	**Lipofectamine mix**
3 µl pPAReni DNA (100 ng/µl)	1.4 µl Lipofectamine 2000
3 µl pFireV5 DNA (100 ng/µl)	129.6 µl Opti-MEM
125 µl Opti-MEM	

3. Both mixes were incubated separately for 5 min at 25°C, pooled and mixed by flicking carefully.

4. The combined master mix was incubated for 25 min at 25°C.

5. 37.4 µl master mix were added to each of the six wells.

6. The cells were incubated at 37°C in a 5% CO_2 atmosphere until the next day.

7. 37.5 µl DMEM were added to each well, containing 0 or 150 ng doxycyclin hyclate, bringing the concentration to 0 or 1 µg/ml, for samples without or with ABL2 induction, respectively.

8. The cells were incubated at 37°C in a 5% CO_2 atmosphere until the next day.

9. The supernatant was removed and the cells were washed with 100 µl PBS per well.

10. Lysis buffer was prepared by adding

500	µl	1 M	HEPES / NaOH pH 7.4
1.5	ml	1 M	NaCl
300	µl	33 mM	EDTA / NaOH pH 8.0
2	ml	50%	glycerol
500	µl	20%	Triton-X 100
100	µl	100%	PTP-Inhibitor mix I
100	µl	100%	PTP-Inhibitor mix II
1		cOmplete Mini protease inhibitor cocktail tablet (Roche)	
5	ml		H_2O

11. Cells were lysed by adding 100 µl lysis buffer to each well and slowly shaking in a nutating fashion on ice for 30 min.

12. From each well, 5 µl of supernatant were taken off, diluted eight-fold by adding 35 µl PBS and stored on ice (input samples).

13. Another 70 µl of supernatant were transferred to an IgG-coated microplate and incubated shaking in a nutating fashion at 4°C for 90 min.

14. The supernatant was discarded and the bound complexes were washed three times with 100 µl PBS per well, each. Finally, 40 µl PBS were added to each well (output samples).

15. 40 µl Bright-Glo Luciferase Assay Reagent (Promega) were added to each sample (input & output) and incubated at 25°C for 5 min shaking in an orbital fashion.

16. Total sample luminescence was measured in a Beckman Coulter DTX 880 integrated over 1 second.

17. Background luminescence was measured in at least six empty wells.

18. The relative multiplicative average for empty wells was calculated by taking the arithmetic mean of the natural logarithm of all empty well measurement.

19. The relative multiplicative standard deviation for empty wells was calculated by adding the squares of the differences between the natural logarithm of each individual empty well measurement and the relative multiplicative average for empty wells, taking the square root of the sum and dividing by the number of measurements minus one.

20. For each condition, i.e. each combination of Protein A-tagged and firefly luciferase-tagged constructs with or without induction, a relative multiplicative average was calculated by taking the arithmetic mean of the natural logarithm of the individual measurements. These values were corrected by taking the logarithm of the antilogarithm of the relative multiplicative average minus the antilorithm of the relative multiplicative average for empty wells, essentially the same as subtracting the empty well measurements before applying the multiplicative model.

21. For each condition, the relative multiplicative standard deviation was calculated by adding the squares of the differences between the natural logarithm of each individual measurement and the relative multiplicative average, taking the square root of the sum and dividing by the number of measurements minus one, i.e. two. These values were corrected by adding the relative multiplicative standard deviation for empty wells.

22. Finally, fold-binding values are obtained taking the antilogarithm of the difference between the condition of interest and the respective comparison, i.e. non-binding protein pairs involving the same firefly luciferase-tagged construct, or, for mutation analyses, the respective wildtype construct.

23. The corresponding Z-scores are calculated by dividing the difference between the condition of interest and the respective comparison by the sum of the relative multiplicative standard deviations.

24. Differences of a factor of at least two and with a Z-score of at least two were considered significant and substantial.

2.7.1 Tyrosine to Phenylalanine Mutations

To test phosphotyrosine-dependent protein-protein interactions in mammalian cell culture, mutant versions of selected ORFS were generated by replacing

tyrosine with phenylalanine residues. These constructs are referenced as "*ORF* Y*(XX)X*F", where *ORF* indicates the template ORF and *(XX)X* indicates the position of the modified tyrosine residue, and were generated by Nouhad Benlasfer in a conservative fashion, i.e. changing TAT and TAC codons to TTT and TTC, respectively.

2.7.2 SH2 Domain Mutations

To test phosphotyrosine-dependent protein-protein interactions in mammalian cell culture, SH2 domain constructs incapable of binding protein tyrosine phosphorylation sites were generated by replacing the most conserved amino acid residue, an arginine canonically making direct contact to the phospho group, with a lysine or a leucine residue. These constructs are referenced as "*ORF* R*(XX)X*K" or "*ORF* R*(XX)X*L", respectively, where *ORF* indicates the template ORF and *(XX)X* indicates the position of the modified arginine residue. The following constructs were generated by Nouhad Benlasfer to make phosphotyrosine dependency accessible in mammalian cell culture:

ORF	**wildtype codon(s)**	**mutant codon(s)**
BCAR3 (R177L)	CGT	CTT
GRB2 (ORF #2 R86K)	CGA	AAG
PIK3R1 (R88KR379K)	CGA CGG	AAG, AAG
PIK3R3 (ORF #1 R90LR383L)	CGA, CGT	CTA, CTT
PIK3R3 (ORF #2 R90LR383L)	CGA, CGT	CTA, CTT
SH2D2A (R120K)	CGG	AAG
SOCS4 (R311K)	CGA	AAG
SOCS4 (R311L)	CGA	CTA

2.8 Analysis of sub-cellular Localization by Immunofluorescence

A subset of protein-protein interactions were visualized in protein fragment complementation assays, determining their subcellular localization. We provided a context for the protein fragment complementation assay results by immunostaining the endogenous proteins, where a specific antibody was available, and overexpressed Protein A-tagged constructs. Confocal images or the mounted samples were produced in a Zeiss LSM700 laser scanning microscope and processed using the accompanying AxioVision software, Version 4.8.2.0 (Carl Zeiss Microimaging GmbH). For each image shown, the transfection, staining and mounting steps were performed by Nouhad Benlasfer.

2.8.1 Coating of glass cover slips with poly-D-lysine

First, glass cover slips were coated with poly-D-lysine to provide a suitable surface for attachment of mammalian HEK293 cells:

1. For each combination of YFP fragment-tagged constructs, a sterile round glass cover slip was placed in one well of a 24-well cell culture plate.

2. To each of these wells, 500 µl poly-D-lysine solution (0.1 g/l in H_2O) were added and incubated at 25°C for 30 min while shaking in an undulating fashion.

3. The supernatant was discarded, the cover slips were washed three times with 500 µl H_2O and completely dried by leaving them under the cell culture hood for 1 hour.

2.8.2 Protein fragment complementation assay

For visualization and determination of subcellular localization of protein-protein interactions in intact mammalian cells, split-YFP fusion constructs were generated by Gateway cloning, co-expressed in HEK293 cells and analyzed by confocal laser microscopy.
Each interaction assessed was assayed in all possible sensible configurations, i.e.,

for ORFs without stop codons at the 3′-end, each interacting protein was fused to both, the N-terminus and the C-terminus, of both, the N-terminal and the C-terminal part of YFP, and assayed in combination with each complimentary construct of the respective interaction partner. ORFs containing a 3′-stop codon were tested in the C-terminal position only. Each fusion construct was further assayed with at least one complimentary construct of a non-interacting protein as a negative control.
Each YFP fragment complementation experiment was considered positive if YFP fluorescence could be observed outside the endoplasmic reticulum. YFP fluorescence in the endoplasmic reticulum only was assumed to result from unspecific adherence of misfolded proteins. To ensure specificity of the signals, only experiments involving only constructs with negative results in other pairings were considered. Positive experiments were documented as digital images.
Detailed protocol:

1. Inoculation was begun by plating one well of 5 x 10^4 T-Rex cells in 500 µl DMEM in 24-well cell culture plates containing a poly-D-lysine-coated glass cover slide in each well for each combination of YFP fragment-tagged constructs, and incubating at 37°C in a 5% CO_2 atmosphere until the next day.

2. For each combination of constructs, a transfection mix was prepared from two pre-mixes:

DNA mix	**Lipofectamine mix**
3 µl pVEN-F1C/F1N DNA (100 ng/µl)	1.4 µl Lipofectamine 2000
3 µl pVEN-F2C/F2N DNA (100 ng/µl)	129.6 µl Opti-MEM
125 µl Opti-MEM	

3. Both mixes were incubated separately for 5 min at 25°C, pooled and mixed by carefully aspirating and dispensing three times with a pipette.

4. The combined transfection mix was incubated for 25 min at 25°C.

5. 250 µl transfection mix were added to the well.

6. The cells were incubated at 37°C in a 5% CO_2 atmosphere until the next day.

7. Cells were fixed by replacing the medium with 500 µl 2% PFA in PBS to each well and incubating at 25°C for 20 min.

8. The PFA solution was discarded. Residual PFA and DMEM were removed by washing twice with 500 µl PBS per well.

9. 500 µl 70% ethanol were added to each well. Samples were stored at -20°C over night or for several days.

10. To stain the cytoskeleton the supernatant was replaced by 200 µl PBS containing 1 U Rhodamine Phalloidin (Invitrogen) and incubating for 20 min at 25°C.

11. The supernatant was removed, the cells were washed twice with 500 µl PBS per well and 100 µl PBS were added to each well.

12. The nuclei were stained by adding 50 µl DAPI (1.6 mg/l in PBS) to each well and incubating for 1 min at 25°C.

13. The cells were washed twice with 500 µl PBS per well.

14. 500 µl 70% ethanol were added to each well and incubated for 1 min at 25°C.

15. The supernatant was replaced with 500 µl 96% ethanol and incubated for 1 min at 25°C.

16. The supernatant was replaced with 500 µl NEO-CLEAR (Merck) and incubated for 1 min at 25°C.

17. For each well, a drop of Depex (Serva) was placed on a glass microscope slide, the cover slide was removed from the 24-well plate and carefully placed on the drop of Depex, face down.

18. Finally, the microscope slides were turned around once more and allowed to try over night.

2.8.3 Detection of endogenous proteins

The endogenous proteins were stained following this protocol:

1. Inoculation was begun by plating one well of 5 x 10^4 T-Rex cells in 500 µl DMEM in 24-well cell culture plates containing a poly-D-lysine-coated glass cover slide in each well for each protein of interest, and incubating at 37°C in a 5% CO_2 atmosphere for two days.

2. The medium was removed. Cells were fixed by adding 500 µl 2% PFA in PBS to each well and incubating at 25°C for 20 min.

3. The PFA solution was discarded. Residual PFA and DMEM were removed by washing twice with 500 µl PBS per well.

4. The cells were permeabilized by adding 500 µl 0.2% Triton-X 100 in PBS and incubating at 25°C for 10 min.

5. The supernatant was removed and the cells were washed twice with 500 µl PBS per well.

6. Unspecific protein binding sites were saturated by adding 150 µl 1% BSA in PBS to each well and incubating at 25°C for 30 min shaking in an undulating fashion.

7. The supernatant was removed and the cells were washed twice with 500 µl PBS per well.

8. The cells were incubated with 200 µl primary antibody solution at 25°C until the next day, shaking in an undulating fashion. The primary antibody solution was prepared by diluting the respective primary antibody in PBS with 0.003% Triton-X 100, diluted 1:100, 1:200 or 1:400 for the GRB2-, RB1- or TSPAN2-specific antibody, respectively.

9. The supernatant was removed and the cells were washed twice with 500 µl PBS per well.

10. The cells were incubated with 200 µl secondary antibody solution at 25°C for 1 h, shaking in an undulating fashion. The secondary antibody so-

lution was prepared by diluting the secondary antibody in PBS with 0.003% Triton-X 100, 1:500.

11. The supernatant was discarded and the cells were washed twice with 500 µl 0.003% Triton-X 100 in PBS per well.
12. 100 µl 0.003% Triton-X 100 in PBS were added to each well.
13. The cytoskeleton was stained by adding 200 µl PBS containing 1 U Rhodamine Phalloidin (Invitrogen) to the supernatant and incubating for 20 min at 25°C.
14. The supernatant was removed, the cells were washed twice with 500 µl PBS per well and 100 µl PBS were added to each well.
15. The nuclei were stained by adding 50 µl DAPI (1.6 mg/l in PBS) to each well and incubating for 1 min at 25°C.
16. The cells were washed twice with 500 µl PBS per well.
17. 500 µl 70% ethanol were added to each well and incubated for 1 min at 25°C.
18. The supernatant was replaced with 500 µl 96% ethanol and incubated for 1 min at 25°C.
19. The supernatant was replaced with 500 µl NEO-CLEAR (Merck) and incubated for 1 min at 25°C.
20. For each well, a drop of Depex (Serva) was placed on a glass microscope slide, the cover slide was removed from the 24-well plate and carefully placed on the drop of Depex, face down.
21. Finally, the microscope slides were turned around once more and allowed to try over night.

2.8.4 Detection of over-expressed Protein A-tagged proteins

To obtain subcellular localization data of proteins for genes without commercially available antibodies and to gauge the influence of overexpression,

ORFs used for protein complentation were overexpressed as Protein A-tagged constructs:

1. Inoculation was begun by plating one well of 5 x 10^4 T-Rex cells in 500 µl DMEM in 24-well cell culture plates containing a poly-D-lysine-coated glass cover slide in each well for each protein of interest, and incubating at 37°C in a 5% CO_2 atmosphere until the next day.

2. For each combination of constructs, a transfection mix was prepared from two pre-mixes:

DNA mix	**Lipofectamine mix**
6 µl pPAReni DNA (100 ng/µl)	1.4 µl Lipofectamine 2000
125 µl Opti-MEM	129.6 µl Opti-MEM

3. Both mixes were incubated separately for 5 min at 25°C, pooled and mixed by carefully aspirating and dispensing three times with a pipette.

4. The combined transfection mix was incubated for 25 min at 25°C.

5. 250 µl transfection mix were added to the well.

6. The cells were incubated at 37°C in a 5% CO_2 atmosphere until the next day.

7. Cells were fixed by replacing the medium with 500 µl 2% PFA in PBS to each well and incubating at 25°C for 20 min.

8. The PFA solution was discarded. Residual PFA and DMEM were removed by washing twice with 500 µl PBS per well.

9. The cells were permeabilized by adding 500 µl 0.2% Triton-X 100 in PBS and incubating at 25°C for 10 min.

10. The supernatant was removed and the cells were washed twice with 500 µl PBS per well.

11. Unspecific protein binding sites were saturated by adding 150 µl 1% BSA in PBS to each well and incubating at 25°C for 30 min shaking in an undulating fashion.

12. The supernatant was removed and the cells were washed twice with 500 µl PBS per well.

13. The cells were incubated with 200 µ l antibody solution at 25°C for 1 h, shaking in an undulating fashion. The antibody solution was prepared by diluting the DyLight 488-conjugated AffiniPure Rabbit Anti-Mouse IgG (H+L) and the DyLight 488-conjugated AffiniPure Donkey Anti-Rabbit IgG (H+L) in PBS with 0.003% Triton-X 100, 1:500 each.

14. The supernatant was discarded and the cells were washed twice with 500 µl 0.003% Triton-X 100 in PBS per well.

15. 100 µl 0.003% Triton-X 100 in PBS were added to each well.

16. The cytoskeleton was stained by adding 200 µl PBS containing 1 U Rhodamine Phalloidin (Invitrogen) to the supernatant and incubating for 20 min at 25°C.

17. The supernatant was removed, the cells were washed twice with 500 µl PBS per well and 100 µl PBS were added to each well.

18. The nuclei were stained by adding 50 µl DAPI (1.6 mg/l in PBS) to each well and incubating for 1 min at 25°C.

19. The cells were washed twice with 500 µl PBS per well.

20. 500 µl 70% ethanol were added to each well and incubated for 1 min at 25°C.

21. The supernatant was replaced with 500 µl 96% ethanol and incubated for 1 min at 25°C.

22. The supernatant was replaced with 500 µl NEO-CLEAR (Merck) and incubated for 1 min at 25°C.

23. For each well, a drop of Depex (Serva) was placed on a glass microscope slide, the cover slide was removed from the 24-well plate and carefully placed on the drop of Depex, face down.

24. Finally, the microscope slides were turned around once more and allowed to try over night.

2.9 Selection of phosphotyrosine-binding genes and ORFs

To provide a comprehensive resource for phosphotyrosine-dependent interactome analysis, we defined a set of phosphotyrosine-recognizing domains, used the InterPro database to identify all proteins comprising these domains, mapped the proteins to Entrez Genes and selected at least one ORF for screening, as far as available. For the second round, we used all available ORFs for the previously unsuccessful genes.
In detail, we took these steps:
Compilation of target Entrez GeneIDs:

1. The InterPro database (Release 19) was queried for all UniProt identifiers associated with the InterPro identifiers

InterProID	Short Name	Description
IPR000980	SH2	SH2 motif
IPR002404	Insln_rcpt_S1	Insulin receptor substrate-1, PTB
IPR006020	PTB_PID	Phosphotyrosine interaction region

2. The amino acid sequences of the UniProt identifiers were retrieved from the UniProt database (Release 14).

3. All RefSeq RNA sequences were retrieved from the RefSeq database (Release 31).

4. For each mRNA, the coding sequence was determined according to the annotations in the RefSeq database and the respective amino acid sequence was generated using the `BioPerl` `Bio::Seq` objects `translate()` function and the `Bio::Tools::CodonTable #1` ("Stardard").

5. A BLAST database was prepared from the translated RefSeq RNA coding sequences.

6. The UniProt sequences comprising the phosphotyrosine-recognizing domain-containing proteins were aligned to the RefSeq RNA sequences

using the NCBI standalone BLAST package (Version 2.0) and the respective BioPerl package. The parameters `e` and `F` were set to `10` and `''F''`, respectively, to ensure inclusion of every query in the results.

7. The BLAST results were evaluated manually to assign each UniProt sequence the matching RefSeq RNA sequences.

8. The Entrez GeneIDs corresponding to the RefSeq RNA identifiers matched by the phosphotyrosine-recognizing domain-containing proteins were obtained using the annotations in the RefSeq database.

Selection of ORFs for screening:

9. The insert sequence of each Gateway Entry clone available in our laboratory was to assigned to an Entrez GeneIDs by matching it to the best fitting RefSeq mRNA using the NCBI standalone BLAST package, as before, except that the `blastn` program was used instead of `blastp`.

10. The Gateway Entry clones assigned to Entrez GeneIDs carrying phosphotyrosine-recognizing domains were selected as potential bait candidates for screening.

11. For each Gateway Entry clone in the phosphotyrosine-recognizing subset, the best matching UniProt identifiers were selected from the set defined by the InterPro identifiers IPR000980, IPR002404 and IPR006020 and aligned using the NCBI standalone BLAST package, as before, except that the `blastx` program was used.

12. For each UniProt identifier in the set defined by the InterPro identifiers IPR000980, IPR002404 and IPR006020 and each of the respective identifiers, minimal domain footprints were defined by the highest number of an amino acid residue annotated as a domain start and the lowest number of an amino acid residue annotated as a domain end, for all overlapping annotations for each identifier in the InterPro database.

13. Gateway Entry clones that failed to comprise at least one minimal domain footprint were disqualified from screening for phosphotyrosine-dependent protein-protein interactions.

14. For round one, for each GeneID with at least one Entry clone in the set of potential bait candidates, the Entry clone that best represented the respective gene was manually selected for screening. Where the best candidate was not obvious, two heuristics were applied. Where available, the Entry clone originating from the hORFeome collection was chosen. Where this was not enough, Entry clones with a $3'$-stop codon were preferred.

15. For round two, for each GeneID with at least one Entry clone in the set of potential bait candidates that has not been included in the first round, all clones were selected for screening. In addition, we obtained Entry clones for genes in the target list, for which we did not possess Entry clones already. For target genes without Entry clones in our selection for which no Entry clones were available from external sources, we obtained non-Gateway-compatible clones. In these cases, the annotations provided by the suppliers were used.

2.10 Statistical Network Analysis

Basic statistical network parameters were computed in Cytoscape (Version 3.1.1) using the NetworkAnalyser tool provided therein:

1. Every protein-protein interaction found in this study was normalized by placing the lower GeneID on the left.

2. The set of unique normalized interactions was supplied with information about protein tyrosine phosphorylation-dependency and imported to Cytoscape (Version 3.1.1).

3. Each node in the network was annotated with boolean attributes whether the respective gene was a successful bait genes, whether the gene was found as prey at least once, and whether the gene was included in the cancer gene census list (Futreal et al. 2004).

4. All loops and all nodes not connected to the largest component were removed.

5. The network was analyzed as an undirected network using the Network-Analyser tool.

6. The parameters average clustering coefficient, diameter, radius, centralization, average shortest path length, average number of neighbors and number of nodes were a direct result of the analysis.

7. The degree distribution was exported and used to calculate the line of best fit assuming a log-log linear distribution weighed by experimental data input (see below).

8. The parameters average clustering coefficient, diameter, radius, centralization, average shortest path length, average number of neighbors and number of nodes were determined accordingly for the set of bait, the set of non-bait genes, the set of cancer genes, the set of cancer genes that are also successful bait genes and the set of cancer genes that are also prey genes by selecting the respective nodes and using the option "Analyze Subset of Nodes".

9. The nodes representing the genes PIK3R3, CRK, CBL, SH2D2A, STAT3, SOCS4, GRB2 and APPL1 were removed and the analysis was repeated to obtain the parameters without core nodes.

10. The process was repeated for the phosphotyrosine-dependent protein-protein interactions only.

11. The data sets provided by Stelzl et al. (2005); Rual et al. (2005); Lim et al. (2006); Wong et al. (2007) were analyzed similarly, except that none of the aspects phosphotyrosine dependency, bait, prey or cancer gene status were regarded.

12. The exponent γ of the line of best fit describing the scale-free model was calculated as

$$\gamma = \frac{\sum_{i=1}^{N} \ln(n_i) \ln(k_i) n_i k_i \sum_{i=1}^{N} n_i k_i - \sum_{i=1}^{N} \ln(n_i) n_i k_i \sum_{i=1}^{N} \ln(k_i) n_i k_i}{\sum_{i=1}^{N} (\ln(n_i))^2 n_i k_i - \left(\sum_{i=1}^{N} \ln(n_i) n_i k_i \right)^2}$$

with N, the number of points in the degree distribution; $i \in \{1, \ldots N\}$;

k_i, the degree of the i-th point in the degree distribution; and n_i, the number of nodes for the i-th point in the degree distribution.

13. The intercept was calculated as

$$\alpha = \frac{\sum_{i=1}^{N} ln(n_i) ln(k_i) n_i k_i - \gamma \sum_{i=1}^{N} ln(k_i) n_i k_i}{\sum_{i=1}^{N} n_i k_i}.$$

14. The equation describes the line of best fit assuming a scale-free model and weighing each data point by the amount of protein-protein interactions involved.

2.11 Alignment of kinase sequence and structure

To visualize the high degree of conservation and position of the residue mutated in generating the catalytically inactive kinase versions, the kinase domains have been aligned on primary sequence and three-dimensional protein structure model levels:
Sequence alignment:

1. The amino acid sequence of the nine kinases were obtained from the UniProt database (Release 14.0) using the previously determined identifiers (P42685, P06241, P51813, P42684, P07332, P43405, Q05397, Q13470, O60674).

2. The kinase domain sequences were obtained by trimming these to the region annotated as the kinase domain in the UniProt database.

3. The sequence alignment was produced with the ClustalW software (Version 2.0.10), hosted by the European Bioinformatics Institute at `http://www.ebi.ac.uk/Tools/msa/clustalw2/`. The alignment options were `protein, slow, gonnet, gap open penalty = 10, gap extension penalty = 0.1` for the pairwise alignment and `gonnet, gap open penalty = 10, gap extension penalty = 0.2, minimum gap distance = 5, clustering method = neighbor joining without iteration` for the multiple alignment.

Structural alignment:

4. The crystal structure models for the kinases were obtained from the RCSB Protein Data Bank, where available, using the identifiers 2DQ7, 2XA4, 3BKB, 3FQS, 3GVU.

5. The structure models were opened in PyMOL (Version 1.4.1) and aligned using the align function provided therein on the respective FLVRES motifs and standard parameters (`cutoff=2.0, cycles=5, gap=-10.0, extend=-0.5, max_gap=50, object=None, matrix='BLOSUM62', mobile_state=0, target_state=0, quiet=1, reset=0, max_skip=0, transform=1`).

2.12 Clustering of kinase specificity patterns

To avoid introducing any kind of bias stemming from human pattern recognition compulsion, we used a computational method to group interactions by kinase specificity according to the kinase plate assay results:

1. Kinase plate assay results were converted to a bipartite graph of interactions and kinases. Edges were placed connecting any pair of interaction and kinase that produced growth in the kinase plate assay.

2. Kinases with less than two spots were removed from the analysis.

3. The bipartite graph was used compute a similarity matrix. For each pair of interactions, the similarity was computed as

$$s_{ij} = \frac{|N_i \cap N_j|^2}{|N_i| \cdot |N_j|}$$

with s_{ij}, the similarity between interactions i and j; and N_n the number of neighbors of the n-th interaction.

4. The obtained similarity matrix was used as input for clustering with affinity propagation clustering algorithm, implemented in R as the APCluster package. The input preference was set to the median for all interactions (`q=0`).

2.13 Semantic similarity of interaction partners

1. The network of phosphotyrosine-dependent interactions, the complete network, as well as the reference networks provided by (Stelzl et al. 2005) and (Rual et al. 2005) were converted to Entrez GeneID and cleared of loops (i.e. self-interactions).

2. A randomized version of the network was generated by adding both GeneIDs to the pool of GeneIDs, for each interaction in the original network. Two GeneIDs were drawn at random. If they were different and the combination was not present in the randomized network, they were removed from the pool and the interaction was added to the network. Otherwise, a random interaction was removed from the nascent randomized network (unless empty) and the two GeneIDs added to the pool. This process was repeated until the pool of GeneIDs was empty. At this point, the randomized network was finished.

3. For each network, 1000 randomized versions were generated this way.

4. For each of the three ontologies (Biological Pathway, Cellular Component, Molecular Function), the 4004 networks (4 input + 4 x 1000 randomized versions) were reduced by the interactions not annotated in the respective ontology.

5. For each interaction in any of the 12012 networks (4004 x 3 ontologies), the semantic similarity of the two GeneIDs according to the respective ontology was calculated using the Bioconductor R package GOSemSim (Version 1.22.0) and the `geneSim` function provided therein. The respective ontology was chosen as `ontology`. Otherwise, the parameters `measure = ''Resnik'', drop = ''IEA'', organism = ''human'', combine = ''BMA''` were used.

6. For each network, the average semantic similarity was calculated by adding all the individual semantic similarity values belonging to the network and dividing by the number of interactions. Interactions for which the semantic similarity could not be computed (”NA”) were valued as `0`, interactions

with an infinite calculated semantic similarity value (''Inf'') were valued as 1.

7. For each of the four network categories, the density of the distribution of the average semantic similarity values was plotted. The average semantic similarity value calculated for the respective input network was indicated by a vertical arrow. The density was calculated using the stats::density() function provided in R using the parameters n=2048, bw=''SJ''.

2.14 Over-representation of GO terms, biological pathways, NESTs and cancer genes

The genes found as prey were tested for statistical over-representation. In other words, the probability of selecting at least the obtained number of genes with a certain annotation from a defined set of genes assuming equal chances of selection for each gene in the set was calculated. A low probability suggests that the selection of genes was biased. To make the obtained probability values more meaningful and more comparable, the values were corrected by multiplying with number of annotations tested. For comparison, the genes used as bait were tested the same way. For GO terms, biological pathways annotated in pathway databases and NESTs, the webserver provided by the ConsensusPathDB was used. The over-representation analysis for cancer genes was calculated using R:

1. The ORFs used as bait were converted to Entrez GeneID as described above. The obtained prey ORFs and the ORFs included in the prey matrix and non-autoactive were converted the same way, if they were derived from Gateway Entry clones. For ORFs derives from cDNA, obtained by classical ligation cloning, partial sequences were available only. The partial sequences were assigned to RefSeq RNA sequences as described above and only the best matching identifier was used for conversion to Entrez GeneID. For each of these categories a unique set of GeneIDs was assembled.

2. For GO term over-representation analysis, the set of prey GeneIDs was uploaded to the ConsensusPathDB (version 21) over-representation analysis tool, the set of matrix GeneIDs was provided as background. The option "`gene ontology level 4 categories`" was chosen, the box for "`biological process`" was checked and `0.01` was set as p-value cutoff. This was repeated for "`molecular function`" instead of "`biological process`", and for "`cellular component`".

3. For pathway over-representation analysis, "`pathways as defined by pathway databases`" was selected instead of "`gene ontology level 4 categories`", and the databases PID, Reactome, NetPath, Wikipathways, KEGG, INOH and Biocarta were selected. Minimum overlap with input list and p-value cutoff were set to `2` and `0.01`, respectively.

4. For network neighborhood-based entity sets (NESTs), "1-next neighbors" was checked with minimum set size, minimum connectivity, minimum overlap with input list and p-value cutoff were set to `2`, `0`, `2` and `0.001`, respectively.

5. The over-representation analysis for cancer genes was performed in R using the `stats::phyper()` function. The function parameters `p, m, n, k, lower.tail, log.p` were set to the number of cancer genes in the prey set, the number of cancer genes in the prey matrix, the number of non-cancer genes in the prey matrix, the total number of genes in the prey set, `FALSE` and `FALSE`, respectively. For interaction-based analyses, the numbers of genes were replaced by the respective numbers of interactions that were found or could have been found. This analyses were carried out twice, once for the complete protein interaction set and once for the phosphotyrosine-dependent interactions, only.

2.15 Linear peptide motifs

Although the focus of this study lies on protein-protein interactions, and no binding surfaces were determined, we estimated how well previously reported linear interaction motif preferences are reflected in our data. For each phosphotyrosine-dependent interaction partner of a phosphotyrosine-recognizing

domain-containing protein with a previously reported motif, we determined the best-fitting tyrosine-centered peptide motif and whether it has been reported to be phosphorylated. For comparison, we did the same for independent interaction partners. We further determined additional parameters for the two data set to refute alternative hypotheses, like the average total number of tyrosine residues and the average total number of reported tyrosine phosphorylation sites, for both sets.
In detail, data were processed according to this protocol:

1. For all protein tyrosine phosphorylation sites reported in the RefSeq (Release 14), UniProt (Release 2010_12) and PhosphoSite (downloaded on October 10, 2010) for human (*Homo sapiens*) or mouse (*M. musculus*), 15-amino acid motifs centered on the reported position were extracted from the respective sequences.

2. Motifs not containing a tyrosine residue in the central position were considered incorrectly annotated and removed.

3. Phosphotyrosine motifs reported in mouse were converted to human, by aligning the identifiers to the respective human identifiers using BLAST, i.e. mouse RefSeq to human RefSeq and mouse UniProt to human UniProt. Phosphotyrosine sites inside aligned regions were converted to human by identifying the best-fitting 15-amino acid residue window for the mouse motif in the human sequence. Human sequence motifs not containing a central tyrosine residue were discarded, as were motifs lying in regions which could not be aligned to a corresponding human sequence.

4. For each ORF that interacted with a phosphotyrosine-recognizing domain-containing protein, the phosphotyrosine motifs reported for the respective identifiers were aligned to the ORFs amino acid sequence. Motifs not contained in the ORF sequence or for which the corresponding ORF sequence was lacking the central tyrosine residue were discarded, the others were considered reported phosphotyrosine sites.

5. For each ORF that interacted with a phosphotyrosine-recognizing domain-containing protein, for each tyrosine residue in the amino acid sequence a 15-amino acid motif was extracted centered on the tyrosine residue.

6. The literature SH2 domain binding motifs aggregated in Table S3 of (Liu et al. 2006) were compared to the corresponding tyrosine-centered motifs by counting the number of matching positions. For each interaction partner, the highest numbers of matching residues was assumed to identify the most likely binding site. The significance of the difference of the distributions for the phosphotyrosine-dependent and the independent interaction partners was assessed by a Wilcoxon rank sum test with continuity correction using the `stats::wilcox.test()` function provided in R and the two distribution vectors as input.

7. For each interaction partner, all best-matching motifs were compared to the reported phosphotyrosine sites and the percentage of reported phosphotyrosine sites was determined by dividing the number of best-matching sites reported as phosphorylated by the total number of best-matching sites, for each category.

8. The average number of tyrosine residues per interaction partner was determined by adding the number of tyrosine residues in all interaction partners, then dividing by the total number interaction partners.

9. The average number of phosphorylated tyrosine residues per interaction partner was determined by adding the number of tyrosine residues reported to be phosphorylated in all interaction partners, then dividing by the total number interaction partners.

10. The average percentage of tyrosine residues reported to be phosphorylated per interaction partner was determined by dividing the number of tyrosine residues reported to be phosphorylated by the total number of tyrosine residues for all interaction partners.

11. The quotients of phosphorylated tyrosine residues and the total number of tyrosine residues in the complete sets was determined by dividing the average number of tyrosine residues per interaction partner by the average number of tyrosine residues per interaction partner.

2.16 Comparison to Protein Interaction Literature

In terms of assessing ones results in the context of data produced by a wide variety of methods and researchers, comparison to literature knowledge is often considered a ”gold standard” for high throughput studies. This kind of comparison is especially useful for estimating false-negative rate and novelty. The novelty estimate is critically dependent on a comprehensive literature data set. In addition, the false-negative estimate requires a defined interaction space. Therefore, we queried the protein-protein interaction meta-database ConsensusPathDB (Kamburov et al. 2009) for each interaction found in our study, as well as each combination of bait genes. We also combined the information to create an estimate of the total phosphotyrosine-dependent interactome size.

2.16.1 Assembly of a curated Literature-based interaction set

First, we generated a manually curated literature interaction data set:

1. Every GeneID of a successful bait gene was combined with each other GeneID of a successful bait gene, generating the complete list of possible interactions in the bait-bait space.

2. The union of the complete list of possible interactions in the bait-bait space and every interaction found in this study was formed, generating the target list for literature database search.

3. The protein-protein interaction meta-database ConsensusPathDB (version 19) was queried for each interaction on the target list. The PubMed identifiers (PMIDs) for each interaction in the results were exported.

4. For each interaction with at least one reported PMID in the ConsensusPathDB, we used the PMID to obtain the publications supposedly reporting the interaction, checking each publication until the interaction was either verified or the list of PMIDs for the interaction was depleted. In this case, the interaction was considered annotated incorrectly and

removed from the set of literature interactions. Interactions that could be verified were used for further analysis. For each publication verifying an interaction we also noted whether there were any hints indicating phosphotyrosine dependence or independence. This annotation was very liberal, for example, interaction induced by a known tyrosine kinase stimulant were considered phosphotyrosine dependent.

2.16.2 Estimation of Novelty

Once a comprehensive literature is available, estimating the novelty is fairly straight-forward:

1. The number protein-protein interactions in the overlap of the literature data set and our results was divided by the total number of interactions found in this study to obtain the fraction of known interactions.

2. The fraction of novel interactions is the complement of the fraction of known interactions. Therefore, we obtained the fraction of novel interactions by subtracting the fraction of known interaction from one.

3. The calculation was repeated for the interactions among bait genes and its complement, i.e. the interactions among one bait gene and one non-bait gene.

2.16.3 Estimation of false-negative rate

For the estimation of false negatives, information about possible results is required. Therefore, the analysis was limited to the bait-bait interaction space:

1. Assuming the number false-positive interactions in the literature set is negligible, the number protein-protein interactions in the overlap of the literature data set and our results was divided by the number of interactions among bait genes in the literature set to obtain the fraction of recovered interactions.

2. The fraction of recovered interactions is the complement of the false-negative rate. Therefore, we obtained the false-negative rate by subtracting the fraction of recovered interactions from one.

3. The false-negative rate for the complete data set is approximated as the square root of the false-negative rate for interactions among bait genes, since the latter have been assayed twice.

2.16.4 Estimation of phosphotyrosine interactome size

To estimate the size of the complete phosphotyrosine-dependent interactome, we assumed that the false-negative rate obtained before is reliable and that our approach is comprehensive, i.e. that the number of phosphotyrosine-dependent interaction among genes without phosphotyrosine-recognizing domains is negligible:

1. We calculated the fraction of recovered interactions for the complete interaction space by subtracting the false-negative rate obtained for the complete interaction space.

2. The approximate number of phosphotyrosine-dependent interactions in the complete interaction space was obtained by dividing the number of interactions found by the fraction of recovered interactions.

Chapter 3

Results

3.1 Comprehensive, genome-scale screening produced a dense network of phosphotyrosine-dependent protein-protein interactions

3.1.1 A modified yeast two-hybrid allows finding phosphorylation-dependent interactions

In conventional yeast two-hybrid analysis (Fields & Song 1989), two plasmids carrying complementary transcription factor hybrids of proteins to be tested for interaction are brought together in a yeast cell carrying reporter gene constructs. The reporters are auxotrophic markers allowing growth on selective medium or enzymes whose activity can be measured in simple enzymatic tests. Such yeast two-hybrid systems are incapable of finding interactions that need post-translational protein modification uncommon to yeast. Therefore, a third plasmid encoding a non-receptor tyrosine kinase gene under the control of a Cu^{2+}-inducible CUP1 promoter was employed for the analysis of phosphotyrosine-dependent interactions. For screening, the kinase plasmid was introduced together with the bait plasmid into MATa yeast strains and mated against an array of MATα prey strains. Diploid strains are able to make histidine only if the two hybrid proteins interact. An enzyme that allows easy colorimetric assaying, β-galactosidase, can be used as additional reporter gene. In a proof-of-principle experiment, the phosphotyrosine-dependently interacting proteins

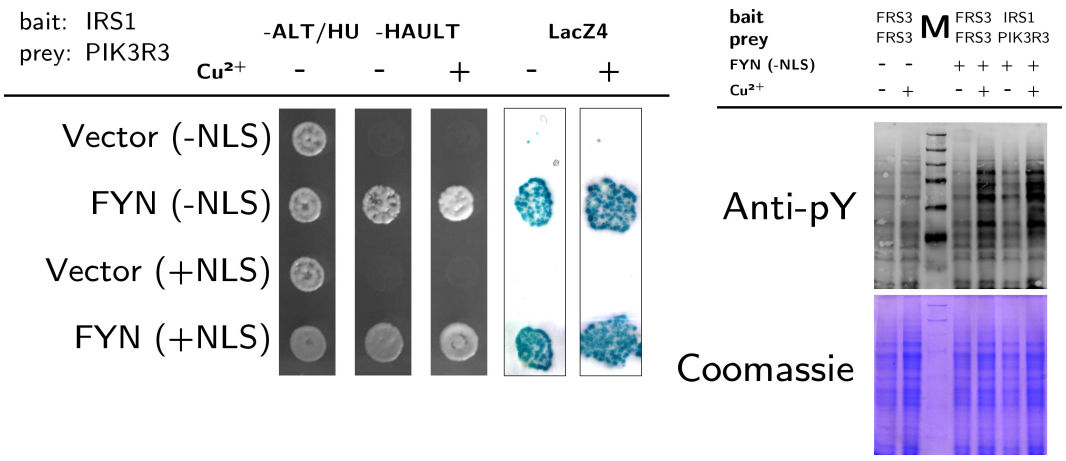


Figure 8: The interaction between IRS1 and PIK3R3 is phosphotyrosine-dependent. (left) Yeast two-hybrid results for the interaction between IRS1 as bait and PIK3R3 as prey in the presence (FYN) or absence (Vector) of a FYN non-receptor tyrosine. The kinase is expressed with (+NLS) and without (-NLS) nuclear localization sequence under the control of a copper-inducible CUP1 promoter. The first column shows a mating control on -ALT/HU medium, the second and third column show growth results on selective -HAULT medium without and with 20 µM copper, respectively . The fourth and fifth column show the result of a β-galactosidase assay corresponding to the second and third column, respectively. The mating control shows growth for all conditions, while the other tests are positive for the kinase-containing yeast only, irrespective of added copper.
(right) Phosphotyrosine-specific immuno blot and coomassie-stained PAA gel of yeast extract. Yeast colonies containing the plasmids for the phosphotyrosine-dependently interacting proteins IRS1 as bait and PIK3R3 as prey and a FYN kinase plasmid without nuclear localization sequence have been grown on selective -HAULT agar without or with 20 µM copper. The yeast proteins were extracted, seperated on a 12% PAA gel, blotted and detected with a phosphotyrosine-specific 4G10 antibody (Anti-pY). For comparison, yeast containing bait and prey plasmids for FRS3, which dimerizes phosphorylation-independently, with and without kinase, was prepared the same way. A second PAA gel that has been run in parallel to the one used for the immuno blot was stained with Coomassie for total protein content and is shown on the bottom (Coomassie), as a loading control.

PIK3R3 and IRS1 (Felder et al. 1993; Pons et al. 1995; Mothe et al. 1997) bestowed the ability to grow on -HAULT medium (Figure 8, top) and produce blue staining in a β-galactosidase assay in the presence of a FYN kinase construct, but not without. When induced, the kinase phosphorylates not only the bait and prey proteins. It also phoshorylates yeast proteins, which can

be seen on Western blots with a phosphotyrosine-specific antibody (Figure 8, bottom). Without induction, the phosphorylation signal on the Western blot is indistinguishable from the control without kinase plasmid, demonstrating the high sensitivity of the growth assay. These results show that introducing an active non-receptor tyrosine kinase into the yeast two-hybrid system makes the detection of phosphotyrosine-dependent protein-protein interactions possible and allows distinguishing kinase dependent PPIs from independent ones in a simple growth assay. Its modular nature, i.e. the fact that it just uses an additional plasmid, makes it very flexible and compatible with a genome-scale prey matrix prepared by Stelzl et al. (2005) and maintained in the Wanker laboratory.

3.1.2 Seventy of 126 genes containing phosphotyrosine-recognizing domains produced interactions in an unbiased, targeted yeast two-hybrid screen

We planned to perform a comprehensive and unbiased screen for phosphotyrosine-dependent interactions using full length proteins. Information about specific amino acid residues, or even proteins, that have been shown to be phosphorylated is far from complete. So, in principle, all human proteins containing at least one tyrosine residue qualify as potentially phosphorylated proteins. The number of phosphotyrosine-recognizing domain-containing proteins, on the other hand, is much better defined. It is also much smaller. It is even smaller than the number of proteins currently known to be tyrosine-phosphorylated. Thus, phosphotyrosine-recognizing domain-containing proteins were screened as bait against a genome-scale prey matrix, under the assumption that prey proteins will be phosphorylated by the coexpressed kinases. This screen will provide a comprehensive approach to a phosphotyrosine-dependent protein-protein interaction network. It is much more feasible and less prone to errors of omission than using e.g. known tyrosine-phosphorylated proteins as bait and screening for phosphotyrosine-recognizing proteins.

There are two major kinds of phosphotyrosine-recognizing domains, SRC homology 2 (SH2) and phosphotyrosine-binding (PTB) domains (Pawson & Scott 1997; Yaffe 2002; Kuriyan & Cowburn 1997). SH2 domains are very clearly

defined and distinguished in terms of structure and function (Liu et al. 2006; Jones et al. 2006; Pawson et al. 2001; Liu et al. 2011a). For PTB domains, there are several subgroups, some of which clearly recognize phosphotyrosine motifs, while others have been shown to bind phospholipids, but do not usually interact with proteins (Rebecchi & Scarlata 1998; Uhlik et al. 2005; DiNitto & Lambright 2006). Structurally, PTB domains have a Pleckstrin homology-like fold and are therefore sometimes classified as a subgroub of the PH domains (Rebecchi & Scarlata 1998; Blomberg et al. 1999). For the purposes of this study, the phosphotyrosine-recognizing domains were defined using the InterPro database (Hunter et al. 2012), focusing on SH2 domains and PTB subgroups which have been widely implicated in phosphotyrosine recognition. UniProt entries containing at least one of the InterPro identifiers IPR000980 (SH2), IPR002404 (PTB_IRS1) or IPR006020 (PTB_PID) were mapped to Entrez GeneIDs using RefSeq sequences (SEE APPENDIX). The three InterPro identifiers were annotated to 418 UniprotIDs, corresponding to 149 Entrez GeneIDs (TABLE X). We planned to obtain at least one full length clone for each GeneID, or, failing that, a clone containing the complete respective domain. All corresponding clones from the Gateway entry clone collection existing in the laboratory, consisting of the hORFeome 1.1 (Rual et al. 2004), 3.1 (Lamesch et al. 2007) and 5.1, supplemented with specific Ultimate ORF clones (Invitrogen) and other clones generated in the context of the NGFN project NeuroNet (BMBF) were used. Additional clones for phosphotyrosine-recognizing genes not found in this collection were obtained from the FLJ entry clone collection (Ota et al. 2004) and the german cDNA consortium (Bechtel et al. 2007). For the genes MAPK8IP1, RIN3, VAV2, PLCG1, APBA2 and STAP2 cDNA clones were purchased and Gateway entry clones were generated by PCR and subsequent Gateway BP shuttling reaction. Altogether, clones for 126 genes (83%) were collected. For 25 genes (17%), no correct clones were obtained (grey entries in Table 4). The inserts of the entry clones were subcloned by Gateway LR reaction into the yeast two-hybrid bait vector pBTM116-D9, prepared for yeast two-hybrid screening (SEE BELOW) and tested for autoactivation. For autoactive GeneIDs, all other clones were tested. For five genes (4% of all tested ones), all available clones were autoactive (yellow rows in Table 4). Non-autoactive strains were screened against the proteome-scale prey matrix. Of

Table 4: The screen covers the target gene set very well. Overview of the target genes, sorted by the Entrez Symbol. For each gene, the Entrez GeneID, the most representative UniProtID, the number of SH2 domains and PTB domains (as defined by the InterPro identifiers IPR000980 (SH2), IPR002404 (IRS1) and IPR006020 (PID)) and the number of clones, is shown. The total number of clones, the number of clones that were screened and the number of successful clones (i.e. clones that produced interactions) are provided. The entries in this column are color-coded according to success state. Genes with at least one successful clone are colored dark green, genes that were screened as bait but failed to produce interactions are colored light green, genes for which all available clones were autoactive (i.e. produced interaction signal in the absence of a prey insert) are colored yellow and genes for which no clones were available are grey.

			Domains			ORFs		
Symbol	GeneID	UniProtID	SH2	IRS1	PID	total	screened	w/PPIs
ABL1	25	ABL1_HUMAN	1			0	0	0
ABL2	27	ABL2_HUMAN	1			1	1	1
ANKS1A	23294	ANKS1_HUMAN			1	1	1	1
ANKS1B	56899	ANS1B_HUMAN			1	2	2	0
APBA1	320	APBA1_HUMAN			1	0	0	0
APBA2	321	APBA2_HUMAN			1	1	1	0
APBA3	9546	APBA3_HUMAN			1	0	0	0
APBB1	322	APBB1_HUMAN			2	2	1	0
APBB2	323	APBB2_HUMAN			2	2	2	0
APBB3	10307	APBB3_HUMAN			2	2	2	1
APPL1	26060	DP13A_HUMAN			1	2	2	2
APPL2	55198	DP13B_HUMAN			1	1	1	1
BCAR3	8412	BCAR3_HUMAN	1			1	1	0
BLK	640	BLK_HUMAN	1			1	1	0
BLNK	29760	BLNK_HUMAN	1			2	1	0
BMX	660	BMX_HUMAN	1			0	0	0
BTK	695	BTK_HUMAN	1			1	1	1
CBL	867	A3KMP8_HUMAN	1			1	1	1
CBLB	868	CBLB_HUMAN	1			2	1	1
CBLC	23624	CBLC_HUMAN	1			2	2	0

Table 4: The screen covers the target gene set very well. (continued)

Symbol	GeneID	UniProtID	Domains			ORFs		
			SH2	IRS1	PID	total	screened	w/PPIs
CCM2	83605	CCM2_HUMAN			1	2	2	1
CHN1	1123	CHIN_HUMAN	1			2	2	0
CHN2	1124	CHIO_HUMAN	1			1	1	1
CISH	1154	CISH_HUMAN	1			1	1	0
CRK	1398	CRK_HUMAN	1			4	3	3
CRKL	1399	CRKL_HUMAN	1			2	1	1
CSK	1445	CSK_HUMAN	1			1	1	1
DAB1	1600	DAB1_HUMAN			1	0	0	0
DAB2	1601	DAB2_HUMAN			1	3	1	0
DAPP1	27071	DAPP1_HUMAN	1			2	2	2
DOK1	1796	DOK1_HUMAN		1		2	1	1
DOK2	9046	DOK2_HUMAN		1		2	1	1
DOK3	79930	DOK3_HUMAN		1		0	0	0
DOK4	55715	DOK4_HUMAN		1		1	1	1
DOK5	55816	DOK5_HUMAN		1		1	1	1
DOK6	220164	DOK6_HUMAN		1		1	1	0
DOK7	285489	DOK7_HUMAN		1		2	1	1
EPS8	2059	EPS8_HUMAN			1	1	1	1
EPS8L2	64787	ES8L2_HUMAN			1	1	1	0
FER	2241	FER_HUMAN	1			1	1	1
FES	2242	FES_HUMAN	1			2	2	0
FGR	2268	FGR_HUMAN	1			1	0	0
FRK	2444	FRK_HUMAN	1			2	2	0
FRS2	10818	FRS2_HUMAN		1		1	1	1
FRS3	10817	FRS3_HUMAN		1		2	2	1
FYN	2534	FYN_HUMAN	1			2	2	1
GRAP	10750	GRAP_HUMAN	1			2	1	0
GRAP2	9402	GRAP2_HUMAN	1			3	1	1

Table 4: The screen covers the target gene set very well. (continued)

			Domains			ORFs		
Symbol	GeneID	UniProtID	SH2	IRS1	PID	total	screened	w/PPIs
GRB2	2885	GRB2_HUMAN	1			3	2	2
GRB7	2886	GRB7_HUMAN	1			2	2	1
GRB10	2887	GRB10_HUMAN	1			1	1	1
GRB14	2888	GRB14_HUMAN	1			1	1	1
GULP1	51454	GULP1_HUMAN			1	0	0	0
HCK	3055	HCK_HUMAN	1			3	3	1
HSH2D	84941	HSH2D_HUMAN	1			1	1	1
INPP5D	3635	SHIP1_HUMAN	1			2	2	0
INPPL1	3636	SHIP2_HUMAN	1			0	0	0
IRS1	3667	IRS1_HUMAN		1		1	1	1
IRS2	8660	IRS2_HUMAN		1		0	0	0
ITGB1BP1	9270	ITBP1_HUMAN			1	2	2	2
ITK	3702	ITK_HUMAN	1			1	1	0
JAK1	3716	JAK1_HUMAN	1			0	0	0
JAK2	3717	JAK2_HUMAN	1			1	0	0
JAK3	3718	JAK3_HUMAN	1			1	1	1
LCK	3932	LCK_HUMAN	1			0	0	0
LCP2	3937	LCP2_HUMAN	1			2	2	0
LDLRAP1	26119	ARH_HUMAN			1	1	1	0
LYN	4067	LYN_HUMAN	1			3	3	1
MAPK8IP1	9479	JIP1_HUMAN			1	1	1	0
MAPK8IP2	23542	JIP2_HUMAN			1	2	2	1
MATK	4145	MATK_HUMAN	1			1	1	0
MIST	116449	CLNK_HUMAN	1			1	1	1
NCK1	4690	NCK1_HUMAN	1			2	2	2
NCK2	8440	NCK2_HUMAN	1			2	0	0
NOS1AP	9722	CAPON_HUMAN			1	1	1	1
NUMB	8650	NUMB_HUMAN			1	1	1	1

Table 4: The screen covers the target gene set very well. (continued)

Symbol	GeneID	UniProtID	Domains			ORFs		
			SH2	IRS1	PID	total	screened	w/PPIs
NUMBL	9253	NUMBL_HUMAN			1	0	0	0
PID1	55022	PCLI1_HUMAN			1	1	1	0
PIK3R1	5295	P85A_HUMAN	2			2	1	0
PIK3R2	5296	Q96CK7_HUMAN	2			1	1	1
PIK3R3	8503	P55G_HUMAN	2			2	2	1
PLCG1	5335	PLCG1_HUMAN	2			1	1	0
PLCG2	5336	PLCG2_HUMAN	2			1	1	1
PTK6	5753	PTK6_HUMAN	1			1	1	0
PTPN11	5781	PTN11_HUMAN	2			1	1	1
PTPN6	5777	PTN6_HUMAN	2			3	3	1
RABGAP1	23637	RBGP1_HUMAN			1	2	2	0
RABGAP1L	9910	RBG1L_HUMAN			1	3	3	1
RASA1	5921	RASA1_HUMAN	2			2	2	1
RGS12	6002	RGS12_HUMAN			1	0	0	0
RIN1	9610	RIN1_HUMAN	1			1	1	1
RIN2	54453	RIN2_HUMAN	1			1	1	0
RIN3	79890	RIN3_HUMAN	1			3	3	3
SH2B1	25970	SH2B1_HUMAN	1			1	1	1
SH2B2	10603	SH2B2_HUMAN	1			0	0	0
SH2B3	10019	SH2B3_HUMAN	1			0	0	0
SH2D1A	4068	SH21A_HUMAN	1			2	2	1
SH2D1B	117157	SH21B_HUMAN	1			3	3	2
SH2D2A	9047	SH22A_HUMAN	1			1	1	1
SH2D3A	10045	SH23A_HUMAN	1			1	1	1
SH2D3C	10044	SH2D3_HUMAN	1			1	1	0
SH2D4A	63898	SH24A_HUMAN	1			1	1	1
SH2D4B	387694	SH24B_HUMAN	1			0	0	0
SH2D5	400745	SH2D5_HUMAN	1			1	1	0

Table 4: The screen covers the target gene set very well. (continued)

			Domains			ORFs		
Symbol	GeneID	UniProtID	SH2	IRS1	PID	total	screened	w/PPIs
SH2D6	284948	SH2D6_HUMAN	1			0	0	0
SH3BP2	6452	3BP2_HUMAN	1			1	1	1
SHB	6461	SHB_HUMAN	1			0	0	0
SHC1	6464	SHC1_HUMAN	1		1	0	0	0
SHC2	25759	SHC2_HUMAN	1		1	0	0	0
SHC3	53358	SHC3_HUMAN	1		1	1	1	0
SHC4	399694	SHC4_HUMAN	1		1	1	1	0
SHD	56961	SHD_HUMAN	1			2	1	1
SHE	126669	SHE_HUMAN	1			1	1	1
SHF	90525	SHF_HUMAN	1			0	0	0
SLA	6503	SLAP1_HUMAN	1			2	1	0
SLA2	84174	SLAP2_HUMAN	1			1	0	0
SOCS1	8651	SOCS1_HUMAN	1			1	1	1
SOCS2	8835	SOCS2_HUMAN	1			1	1	0
SOCS3	9021	SOCS3_HUMAN	1			2	1	1
SOCS4	122809	SOCS4_HUMAN	1			1	1	1
SOCS5	9655	SOCS5_HUMAN	1			2	2	0
SOCS6	9306	SOCS6_HUMAN	2			1	1	1
SOCS7	30837	SOCS7_HUMAN	1			0	0	0
SRC	6714	SRC_HUMAN	1			2	2	2
SRMS	6725	SRMS_HUMAN	1			1	1	0
STAP1	26228	STAP1_HUMAN	1			1	1	0
STAP2	55620	STAP2_HUMAN	1			1	1	0
STAT1	6772	STAT1_HUMAN	1			1	1	1
STAT2	6773	STAT2_HUMAN	1			0	0	0
STAT3	6774	STAT3_HUMAN	1			4	2	2
STAT4	6775	STAT4_HUMAN	1			0	0	0
STAT5A	6776	STA5A_HUMAN	1			2	1	1

Table 4: The screen covers the target gene set very well. (continued)

			Domains			ORFs		
Symbol	GeneID	UniProtID	SH2	IRS1	PID	total	screened	w/PPIs
STAT5B	6777	STA5B_HUMAN	1			2	0	0
STAT6	6778	STAT6_HUMAN	1			1	1	0
SUPT6H	6830	SPT6H_HUMAN	1			0	0	0
SYK	6850	KSYK_HUMAN	2			2	1	1
TBC1D4	9882	TBCD4_HUMAN			2	1	1	0
TEC	7006	TEC_HUMAN	1			1	1	0
TENC1	23371	TENC1_HUMAN	1		1	1	1	1
TNS1	7145	TENS1_HUMAN	1		1	0	0	0
TNS3	64759	TENS3_HUMAN	1		1	1	1	0
TNS4	84951	TENS4_HUMAN	1			1	1	1
TXK	7294	TXK_HUMAN	1			1	1	0
TYK2	7297	TYK2_HUMAN	1			1	1	0
VAV1	7409	VAV_HUMAN	1			1	1	0
VAV2	7410	VAV2_HUMAN	1			1	1	0
VAV3	10451	VAV3_HUMAN	1			1	1	0
YES1	7525	YES_HUMAN	1			1	1	1
ZAP70	7535	ZAP70_HUMAN	2			2	2	0
					total	188	159	82

the 126 genes screened, 70 (56%) were functional, i.e. produced at least one interaction (dark green in Table 4). Fifty-one genes (40%) were screened but failed to produce any interaction as bait (light green in Table 4). On the clone level, 188 phosphotyrosine-recognizing gene-coding sequences were cloned and tested for autoactivity, 159 (85%) clones were screened and 82 (44%) clones were functional. This is in agreement with previous large scale yeast two-hybrid screens, that show a similar fraction of clones that did not result interactions (Stelzl et al. 2005; Rual et al. 2005; Goehler et al. 2004). The relative numbers are smaller for the clones, because more clones were tried for genes that were autoactive or non-functional.

In conclusion, a comprehensive set of phosphotyrosine-recognizing domain-containing genes was screened against a proteome-scale prey matrix. In these experiments, about half of all phosphotyrosine-recognizing domain-containing genes interacted with at least one prey protein.

3.1.3 Genome-scale yeast two-hybrid screening of a comprehensive phosphotyrosine gene set produced 292 phosphotyrosine-dependent interactions

A comprehensive set of phosphotyrosine-recognizing proteins was selected as bait for genome-scale screening for phosphotyrosine-dependent protein-protein interactions. Towards this end, the relevant tyrosine residues in the potentially interacting prey proteins had to be phosphorylated in the yeast cellular environment. There are no endogenous bone fide tyrosine kinases in *S. cerevisiae* (Manning et al. 2002a; Gnad et al. 2009), so they were supplied exogenously. In our experiments, they were supplied on an additional ade2 selectable plasmid under the control of a copper-inducible CUP1 promoter (Butt et al. 1984). We used one member of each non-receptor tyrosine kinase family, except for the CSK family which is known to exclusively add repressive phopsphotyrosine marks to non-receptor tyrosine kinases (Levinson et al. 2008). The use of representative kinases of different families increases the chances of phosphorylating each relevant tyrosine residue in the assay. The yeast two-hybrid MATa strains were successively transformed with the kinase and the bait vectors. The resulting yeast strains were arrayed in twenty sets of two microplates each. Each plate contained up to eight different bait genes arranged in rows and up to nine kinases arranged in columns. Each such set of two microplates constitutes one complete set of kinases and controls for eight bait proteins and was screened as one pool.

The proteome-wide screen was performed in three main steps: test for autoactivation, primary screen and retest. As an example, Figure 9 shows the autoactivation test of pool 11, the primary screen of pool 11 with matrix plate 26 and the corresponding retest. Before the genome-scale screen against the prey matrix, all strains were tested for autoactivation (Figure 9A) to avoid collecting false positives and masking of true positive interactions when using a pooled

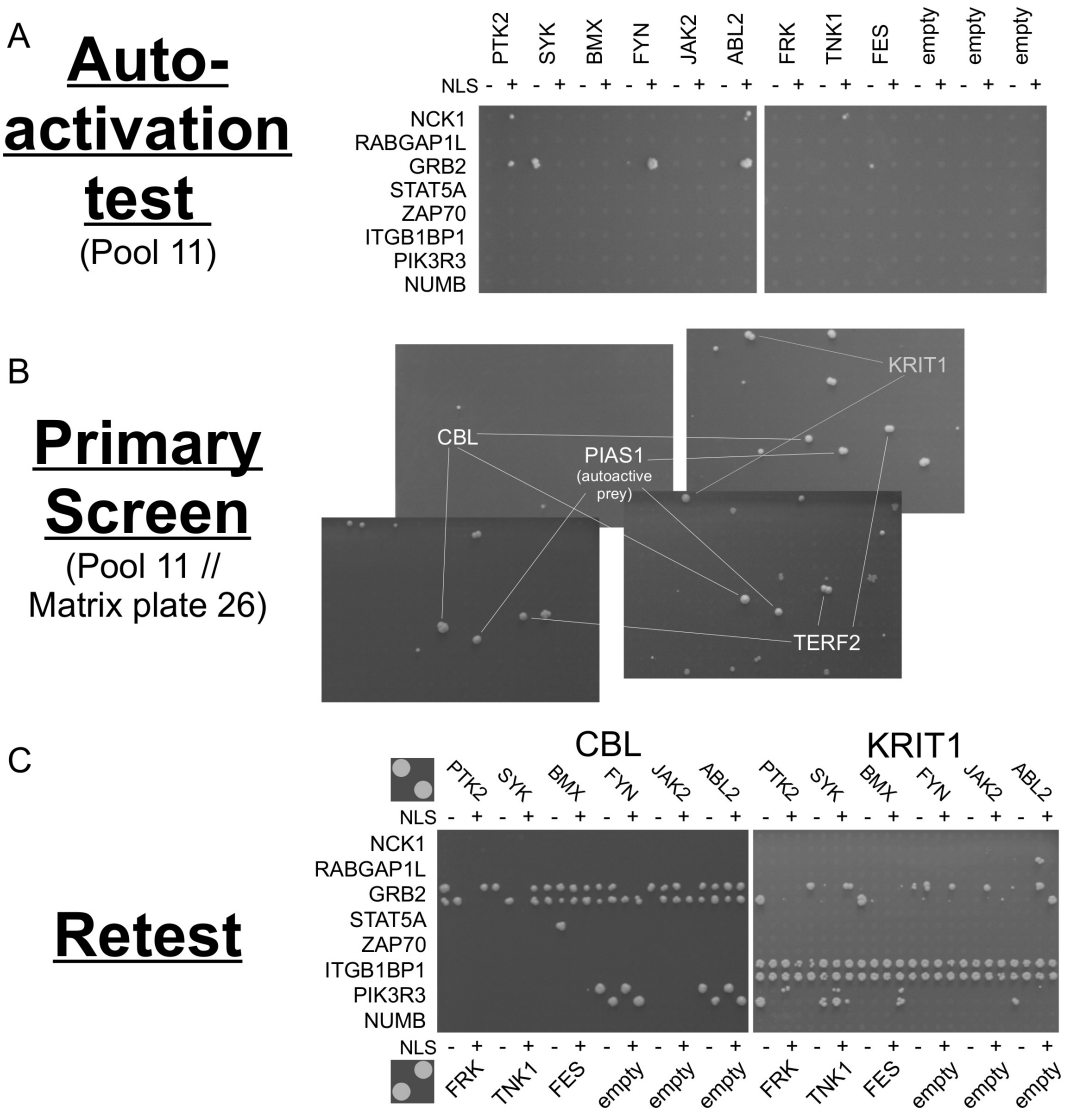


Figure 9: The matrix screen produces high-quality phosphotyrosine-dependent and independent protein interactions. Exemplary illustration of the matrix screening process of autoactivation test, primary screen and retest using bait pool 11 and matrix plate 26. For the autoactivation test, the array of bait strains with kinases were mated with an empty prey vector strain, diploids were recovered on -ALT/HU agar and transferred to -HAULT(Cu^{2+}) agar (A). Any positions growing in the autoactivation test, like the ones seen for NCK1 and GRB2, were removed from the assay. The others were grown separately, mixed and used as bait for a genome-scale matrix screen. This was repeated for four independent replicates. Prey strains growing in at least two of the four repeats (B) that were not known autoactivators (i.e. prey strains growing with almost any bait) were retested using fresh yeast (C). In the retest, the bait strains were arrayed in 384-well format, and tested against single prey strains. The retest plate for CBL (C, left) shows a phosphotyrosine-dependent

approach. To test for autoactivation, the bait strains were mated with a MATα strain containing a prey vector without insert. This procedure provides a diploid yeast environment with an expressed RNA polymerase-recruiting domain. To ensure diploid recovery, after mating, the yeast was first spotted on -ALT/HU agar and controlled for full growth. The -ALT/HU spots were then transferred to -HAULT(Cu^{2+}) agar to test for autoactivation. Any bait strains growing on -HAULT(Cu^{2+}) agar in the autoactivation test, were removed from the pool for screening (like the spots seen for NCK1 and GRB2 in Figure 9A). Strains that did not grow on -ALT/HU agar were also removed from the respective pools or the autoactivation test was repeated. In the primary screen, strains contained in separate positions of the microplates were grown separately in deepwell plates, pooled, mixed and mated against a matrix of individual prey strains. After mating, the diploids were transferred directly to selective agar (-HAULT(Cu^{2+})) and growing colonies were recorded as primary hits (Figure 9B). Each pool was screened this way four times using independent replicas.

In the evaluation of primary hits, known autoactivating prey strains were disregarded. All remaining combinations of bait pools and prey strains that led to yeast growth in at least two out of the four replicate screens were retested independently using fresh yeast. Figure 9B shows four potentially interacting prey genes for pool 11 on matrix plate 26, i.e. four positions growing in at least two replicas. PIAS1 is a known autoactive prey and was excluded from retesting, CBL, TERF2 and KRIT1 were retested. For retesting, the 96-well bait pool microplates were merged into 384-well microplates with two copies of identical combinations of bait constructs and kinases at diagonal, i.e. non-neighboring positions. The prey strains corresponding to the primary hits were mated against the microplate corresponding to the bait pool (Figure 9C). This step not only provided greater stringency by confirming the interaction in a second, independent experiment, but also deconvoluted interacting bait constructs

interaction with PIK3R3 that allows growth only in the presence of an ABL2 or FYN kinase and an independent interaction with GRB2. KRIT1 (C, right) interacted with ITGB1BP1.[a]

[a]Note that all bait clones were controlled after the screen and the NCK1 clone was removed from the data set because it produced questionable results.

and kinases, effectively discriminating phosphotyrosine-dependent interactions from independent ones. It serves as an additional control for prey autoactivity, because only one or two bait constructs in any pool are expected to interact with any given prey. The left side of Figure 9C shows the retest for CBL, the right side for KRIT1. CBL interacted with two of the bait genes in the pool. It interacted with GRB2 in a phosphorylation-independent manner allowing growth on -HAULT(Cu^{2+}) medium irrespective of the presence or identity of the kinase. The interaction between PIK3R3 and CBL, on the other hand, is phosphorylation-dependent. The two proteins interacted in the presence of a FYN or ABL2 kinase only. This resulted in a very characteristic growth pattern of diagonally adjacent spots on -HAULT(Cu^{2+}) agar. The other retest result shows an independent interaction between ITGB1BP1 and KRIT1. The weak growth/few spots seen on the KRIT1 retest plate for RABGAP1L, GRB2 and PIK3R3 and on the CBL retest plate for STAT5A were either autoactives known from the autoactivation test, late autoactives or the interactions were too weak or produced too few spots to be considered.

The yeast two-hybrid screen was performed in two rounds. In the first round

Table 5: The Two-Step Screening Protocol improved the output substantially. Overview of parameters of the yeast two-hybrid screens. The number of bait, kinase and prey genes screened, as well as the size of the covered interaction space is shown for the whole screen (total) and for the two rounds individually (Round 1 & Round 2). Additionally, the size of the retest experiment and the output in term of total, phosphotyrosine-dependent (pY-dependent) and phosphotyrosine-independent (pY-independent) protein-protein interactions is presented.

Number of	Round 1	Round 2	Total
bait genes	80	41	121
kinase genes	9	3	9
prey genes	13,807	17,007	17,007
possible interactions	1,104,560	697,287	1,801,847
double, triple & quadruple hits retested	899	324	1,223
interactions	359	270	628
–pY-dependent	172	121	292
–pY-independent	187	149	336

of screening eighty bait genes were screened in combination with nine kinases (Table 5). In the meantime, bait clones for 41 genes missing in our collections or without interactions (because the first clones was autoactive or unsuccessful) were obtained. These clones were screened in the second round (Table 5). At this point, the number of clones in the prey matrix had been extended from 13,807 to 17,007 in the laboratories of Erich Wanker at the Max Delbrück Center for Molecular Medicine in Berlin. The number of different kinases employed was reduced from nine to three, because the other six returned very few interactions with the first set of bait constructs tested and the workload reduction was substantial. The second round of screening expanded the probed interaction space by about 63%, adding another 697,287 possible interactions to the 1,104,560 possible interactions probed in the first round (Table 5). Consequently, the number of interactions increased by 120 tyrosine phosphorylation-dependent and 149 independent interactions. In total, 121 bait genes were screened with up to nine kinases and up to 17,007 prey strains, covering an interaction space of almost two million possible interactions in four replicate screens. In these screens, 1,223 combinations of bait pools and prey strains came up at least twice and were retested. After retesting, these amounted to 292 phosphotyrosine-dependent and 336 independent interactions.

3.1.4 The phosphotyrosine-dependent interactions form a network of unusually high density

In a genome-scale screen, a comprehensive set of genes involved in phosphotyrosine signaling has been screened for phosphorylation-dependent protein-protein interactions. As a result of this screen, 292 phosphorylation-dependent and 336 independent protein-protein interactions involving 82 bait clones have been established in the retest. The set of bait genes screened covers the whole phylogeny of SH2 domains (Figure 10). The interacting SH2 domain-containing proteins are distributed across the phylogenetic tree reflecting the sequence similarity and evolution of human SH2 domains (Liu et al. 2006). The SH2 domain-containing proteins with the highest number of interaction partners are PIK3R3, CRK, STAT3, TENC1, CLNK, SH2D2A and GRB2. Generally we observed that, if a functional clone exists for a gene, it produces tyrosine

phosphorylation-dependent interactions, as well as independent ones. TENC1 is the only gene with a high number of interactions, but without phosphorylation-dependent ones. Except for the clades of VAV1, VAV2 and VAV3 and SHC1, SHC2, SHC3 and SHC4, most of which have been screened, but did not produce interactions, there is no apparent group of sequence-related genes without work-

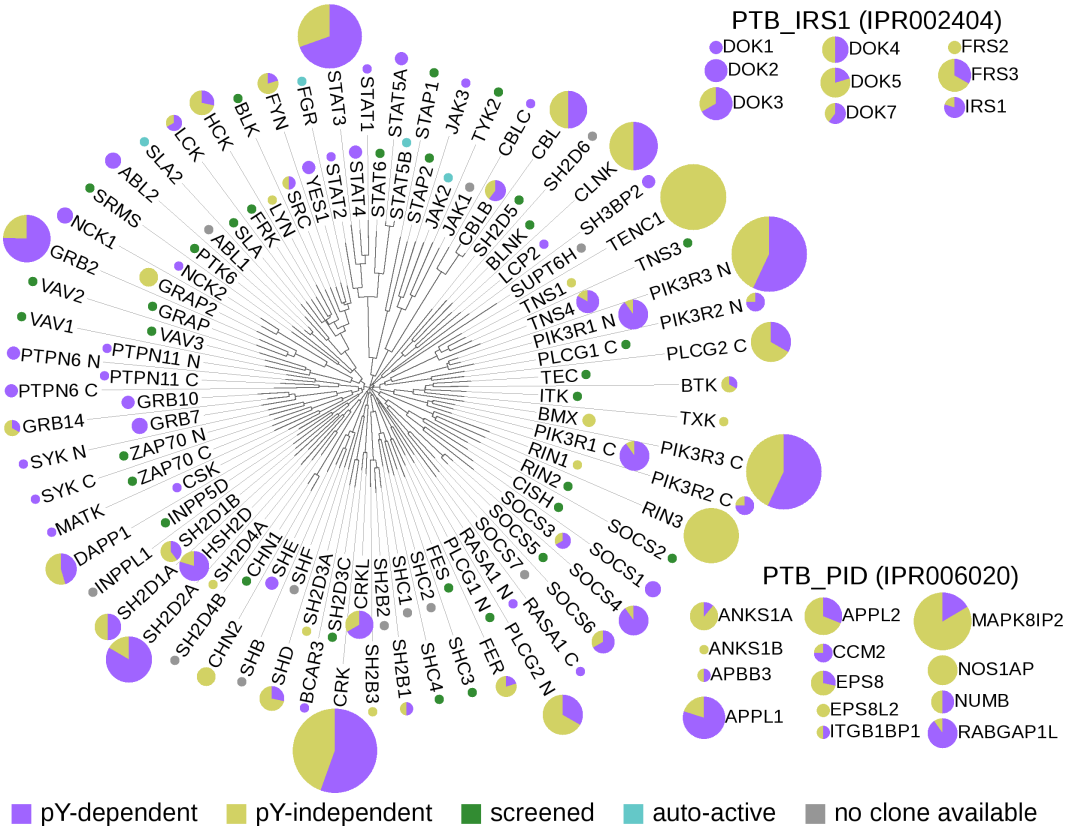


Figure 10: The phosphotyrosine-dependent protein-protein interactions are distributed among the SH2 domain phylogenetic tree. Phylogenetic tree reflecting the evolution of SH2 domains. For each SH2 domain, the number and type of interactions is indicated by the size and color of the pie charts. The phylogenetic tree is based on the SH2 domain peptide structure alignment presented in Liu et al. (2006), which has been manually corrected based on tertiary structure information. The number of interactions is proportional to the area of the pie. Phosphotyrosine-dependent interactions are signified by the purple area, independent interactions by the yellow area. SH2 genes without interactions are coded by circles of size one according to class. Screened as bait but without interactions is green. Genes with autoactive clones (i.e. clones inducing growth regardless of the prey) only are blue, genes without clones are grey. PTB domain-containing without SH2 domains are shown in separate boxes. Only genes with interactions are shown.

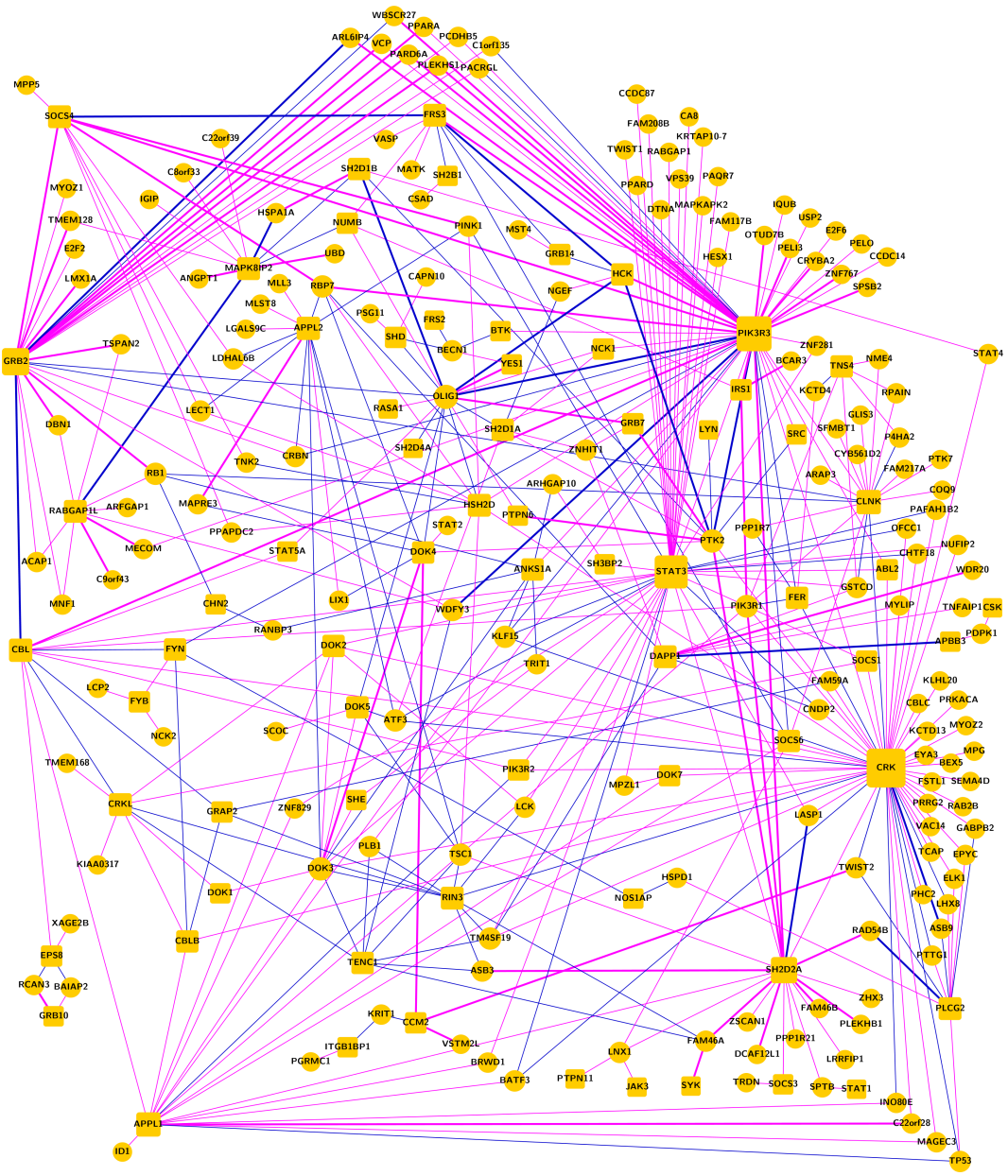


Figure 11: The network of phosphotyrosine-dependent protein-protein interactions is highly connected. "Hairball" representation of the phosphotyrosine-dependent protein-protein interaction network. Proteins are represented by nodes, links represent protein-protein interactions. Phosphotyrosine-independent interactions linking the nodes are shown in blue, phosphotyrosine-dependent interactions are purple. Strong links signify protein interactions validated in mammalian cell culture. The node shape is rectangular with rounded corners for bait genes and round for genes that were screened as prey only.

ing clones. For the PTB domain-containing genes, we observed a trend for genes with the IRS1 subtype to show a higher fraction of phosphorylation-dependent interactions compared to the PID. It has to be noted however that the former are strongly dominated by DOK proteins, which, while not extremely similar in primary sequence, all contain a PH domain in addition to the PTB domain. The latter have more total interactions per gene.

Figure 11 shows a graphical representation of the phosphorylation-dependent protein-protein interaction network. Nodes represent proteins, links represent protein-protein interactions. The phosphotyrosine-dependent interactions are represented by purple links, and independent interactions linking the nodes have been added in blue. Proteins that were screened as bait are represented by rectangular nodes with rounded corners, proteins that interacted as prey only are round. The network appears to be very dense and highly connected. It has several highly connected nodes and many nodes with few interactions only making it seem scale-free. It contains many well-characterized interactions, like the one between PIK3R3 and IRS1 (Felder et al. 1993; Pons et al. 1995; Mothe et al. 1997) or the one between CBL and CRK (Fukazawa et al. 1996), as well as interactions with virtually uncharacterized proteins like C10orf81, which interacted with both, PIK3R3 and GRB2. To substantiate the impressions about the network structure we performed systematic tests.

Figure 12 shows the networks degree distribution. The network satisfies the small world criterion and is scale-free. That is, it contains many nodes of low degree and few nodes, called hubs, of high degree, making the average shortest path length very small (Cohen & Havlin 2003). It has a log-log linear degree distribution, and, hence, no scale (Albert & Barabási 2002). The unusually high density of the network can also be seen in the degree distribution. Scale-free degree distributions can be described by an exponential function. The exponent is expected to lie between -2 and -3, but it is -1.41 for the data in this study (Figure 12) (Yook et al. 2004; Cohen & Havlin 2003). This means that the ratio of high-degree nodes to low-degree nodes is higher than in other protein-protein interaction networks.

The network of phosphorylation-dependent interactions alone has a giant component comprising all but eleven interactions, namely the interactions of APPL2, GRB10, SHD, ITGB1BP1 and SH2B1. Four of these five proteins are linked

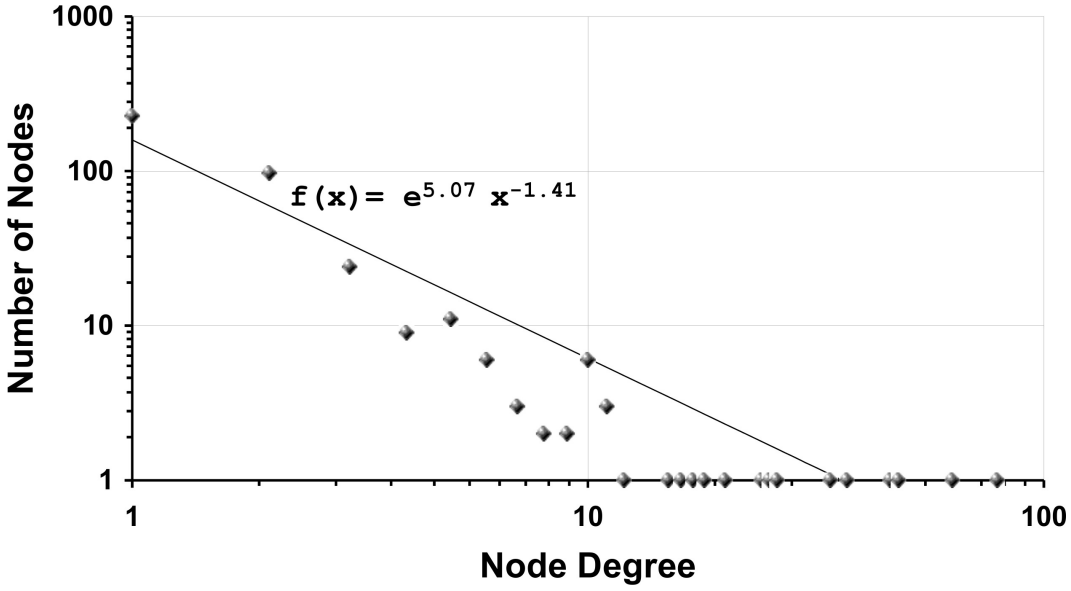


Figure 12: The network of phosphotyrosine-dependent protein-protein interactions is scale-free. Degree distribution of the phosphotyrosine-dependent protein-protein interaction network. The number of nodes for each degree is plotted against the node degree. Both axes have a logarithmic scale. The distribution fits a scale-free distribution described by the formula $f(x) = \mathrm{e}^{5.07} x^{-1.41}$.

to the giant component by a single phosphorylation-independent interaction. Only ITGB1BP1 has a distance of two independent interactions to the giant component. The network has a core formed around the eight genes PIK3R3, CRK, CBL, SH2D2A, STAT3, SOCS4, GRB2 and APPL1 (Table 6). Removal of these core nodes led to drastic losses of clustering coefficient and centrality, accompanied by a substantial increase in average shortest path length, radius and diameter. Taken alone, the network of phosphorylation-dependent interactions broke apart, decreasing the size of the largest component from 201 to 59 nodes. Simultaneously, the average path length, diameter and radius increased. There are 51 phosphorylation-dependent and 30 independent bait-bait interactions. This number is much greater than expected by chance. Given the search space and the total number of interactions found and assuming an average number of interactors per gene (independent of whether the prey is also a member of the bait set), one would expect little more than eight bait-bait interactions in the complete set and less than three in the set of phosphotyrosine-dependent interactions alone. The odds of seeing the much greater numbers we actually

Table 6: Characteristics of the phosphotyrosine-dependent protein interaction network Network parameters of the phosphorylation-dependent protein-protein interaction network with ("complete network") and without ("pY-dependent network") independent interactions. "Bait/Nonbait only": Considering only nodes that are either bait genes or not. "Without core": Statistics for the same network without PIK3R3, CRK, CBL, SH2D2A, STAT3, SOCS4, GRB2 and APPL1.

	Avg. Number of Nodes	Avg. Number of Neighbors	Clustering Coefficient	Centralization	Diameter	Radius	Shortest Path Length
pY-dependent network	201	2.726	0.027	0.219	9	5	3.823
bait only	50	6.500	0.048	n.a.	n.a.	n.a.	n.a.
nonbait only	151	1.477	0.020	n.a.	n.a.	n.a.	n.a.
without core	59	2.237	0.000	0.174	11	7	4.980
complete network	405	3.081	0.026	0.189	9	5	3.751
bait only	69	10.203	0.074	n.a.	n.a.	n.a.	n.a.
nonbait only	336	1.619	0.016	n.a.	n.a.	n.a.	n.a.
without core	270	2.526	0.010	0.170	11	6	4.717

observe by chance are 4.41×10^{-58} and 6.31×10^{-38} for the complete network and the network of phosphorylation-dependent interactions only, respectively. These observations together with the observation that the network is kept together via a small core set of PPIs and the relatively low power coefficient of the degree distribution suggest that phosphotyrosine signaling involves a relatively restricted subset of proteins of the human proteome.

In conclusion, the network shows characteristics expected for a protein-protein interaction network, but is more coherent than seen for other large protein-protein interaction networks (Yook et al. 2004; Cohen & Havlin 2003). It contains a high number of bait-bait interactions, suggesting a strong functional connection (Bader & Hogue 2003; Bu et al. 2003; Spirin & Mirny 2003). These observations are more pronounced for the network of phosphorylation-dependent interactions alone than for the complete network.

3.2 Kinase specificity of the phosphotyrosine-dependent protein-protein interactions in the yeast two-hybrid system is non-trivial

Kinase specificity in the yeast system can be investigated in great detail in a "kinase plate assay". For this assay, pairs of interacting bait and prey constructs are prepared in haploid MATα yeast and mated against an array of MATa yeast carrying different kinase plasmids in microplate format. Each bait-prey interaction was tested in parallel with 31 non-receptor tyrosine kinases in the kinase plate assay. A crucial point is that the yeast is pipetted on Omnitray agar plates. The relatively large amount of 5 µl yeast suspension per spot is more than an order of magnitude greater than the amount spotted by stamping. This provides high confidence not only in positive results, but also in negative results. Differential growth signals with different kinases are specific to the interactions. Because of the added confidence in negative results, the kinase plate assay serves another purpose, as well. In addition to non-receptor tyrosine kinases and a number of empty vector controls, the kinase plate contained kinase-deficient mutant versions of all kinases used for screening. It demonstrates that the interactions that show a differential signal in the retest with the kinase-deficient mutant when compared to the active kinase are indeed phosphorylation-dependent, not just kinase-dependent.

Like all human protein kinases, non-receptor tyrosine kinases have a conserved kinase domain, that is made up by an amino-terminal lobe rich in anti-parallel β-strands responsible for ATP binding and a C-terminal lobe rich in α-helical structures, that contains the activation loop (Figure 13a and b) (Hubbard & Till 2000). The two lobes are connected by a less conserved linker region. The structurally or functionally important residues in the lobe regions are generally highly conserved. Interestingly, the activation loop is not well conserved sequence-wise and very flexible structure-wise (Figure 13a and b, purple). Therefore, a conserved lysine residue in the N-terminal lobe that is necessary for ATP binding (Figure 13a and b, red) was replaced by methionine to abrogate kinase activity (Kamps et al. 1984). The kinase plate contains 31 of 32 described human non-receptor tyrosine kinases (Figure 13c) (Robinson et al. 2000). For JAK1 several

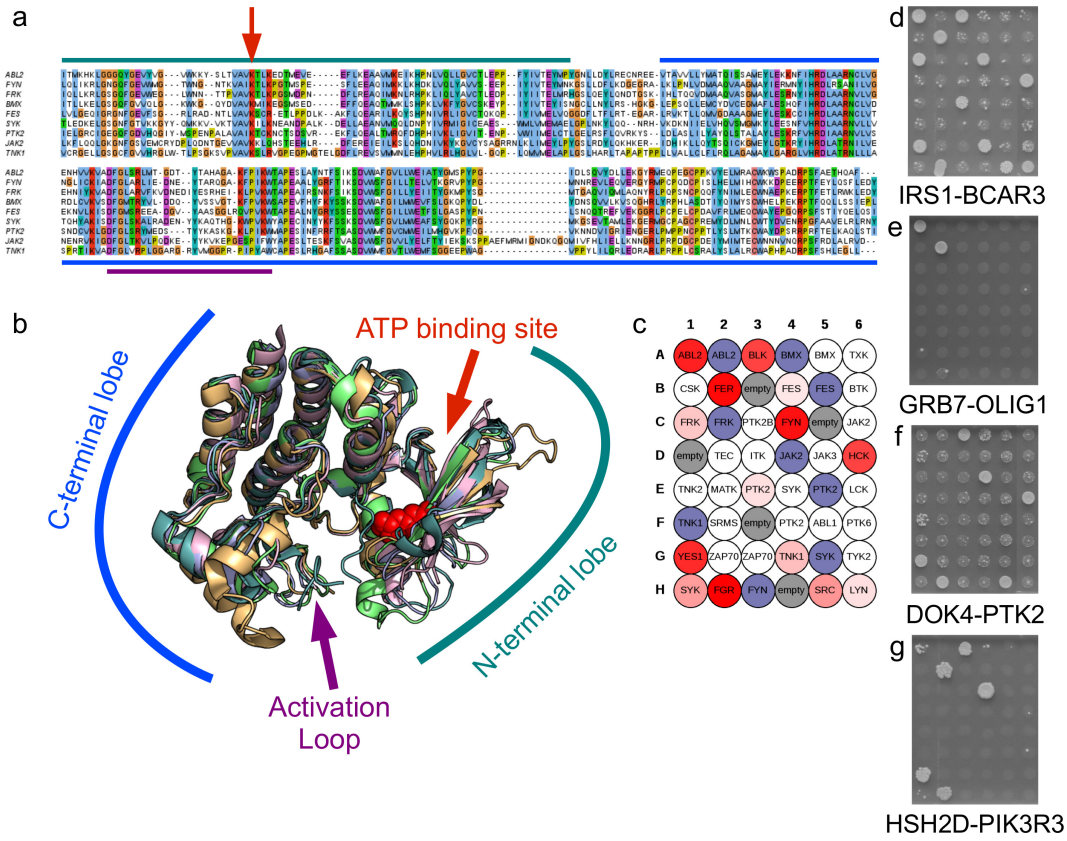


Figure 13: ATP binding-deficient kinases qualify kinase-dependent interactions as phosphotyrosine-dependent. (a) Amino acid sequence alignment of the kinase domains of all kinases used for yeast two-hybrid screening. UniProt entries P42684 (ABL2), P51813 (BMX), P07332 (FES), P42685 (FRK), P06241 (FYN), O60674 (JAK2), Q05397 (PTK2), P43405 (SYK) and Q13470 (TNK1) were used to define kinase domain boundaries. The red arrow points out a conserved lysine residue in the ATP binding cassette that has been shown to disrupt ATP binding when mutated to methionine. (b) Protein structure model alignment of the kinase domains of all kinases used in the screen for which crystal structure models in the PDB. Polypeptide backbone structures are shown as cartoon colored according to kinase, the lysine responsible for ATP binding, that has been mutated to generate the kinase-deficient versions is shown as red spacefill. Light blue: FYN (PDB accession 2DQ7, UniProt accession P06241, residues 271-524), light orange: JAK2 (2XA4, O60674, residues 849-1124), light pink: FES (3BKB, P07332, residues 561-822), dark teal: SYK (3FQS, P43405, residues 371-631), lime: ABL2 (3GVU, P42684, residues 288-539). Protein structure models have been aligned using the PyMol align function. (c) Arrangement of kinases in the kinase plate assay. For active kinases, the degree of redness indicates the activity (i.e. the number of interactions growing with the kinase), kinase-deficient mutants are colored blue and empty vector controls are grey. (e-g) kinase plate assay results on -HAULT(Cu^{2+}) agar the interactions IRS1-BCAR3 (d), GRB7-OLIG1 (e), DOK4-PTK2 (f) and HSH2D-PIK3R3 (g).

clones were obtained and analyzed, but turned out to be falsely annotated, suggesting an annotation error of the MGC clone (Temple et al. 2006). For PTK2, SYK and ZAP70 two clones representing different isoforms were used. Empty kinase vectors serve as controls for late autoactivity and to sort out conceivable interactions that can be bridged by non-receptor tyrosine kinases irrespective of protein phosphorylation activity. The kinases were arranged on the microplate to avoid clusters of spots expected to grow or expected not to grow for phosphorylation-dependent interactions (Figure 13c). Interactions were assayed with kinases in the kinase vectors pASZ-CN-DM and pASZ-C-DMtet, i.e. with and without nuclear localization signal sequence, in two replicates each. Some kinases work equally well in both kinase vectors, while others display a clear difference.

The kinases showed different levels of activity (highly active kinases are shown in red in Figure 13c). The most active kinases (i.e. the kinases that enable the highest number of interactions) were the Src family kinases FYN, HCK, YES1, BLK and FGR and the non-Src family kinases ABL2 and FER. This could be a consequence of the different modes of regulation. Where most other kinases are inactive by default and biochemically activated following an appropriate signaling event, Src family kinases, Abl kinases and FER are active by default and inactivated by CSK (Okada et al. 1991), PRDX1 (Wen & Van Etten 1997) and PLEC (Lunter & Wiche 2002), respectively. It is therefore conceivable that these kinases are the most active ones, because the inhibitory proteins are simply not present in yeast. Interacting pairs of bait and prey constructs function with different numbers of kinases. For example, some interactions, like the one between IRS1 and BCAR3 (Figure 13d) function with twelve kinases, while others, like the one between GRB7 and OLIG1 (Figure 13e) is enabled by only two kinases in our assay. The most frugal explanation for these two facts is that the kinases differ in catalytic activity leading to higher or lower phosphorylation levels and interactions display different levels of sensitivity needing higher or lower phosphorylation levels to enable yeast growth. This would result in patterns that are subsets of each other. This not what we observe. Clearly, the kinase plate assay patterns point towards interaction specificity that can be seen for example in the difference between the kinase patterns of GRB7-OLIG1 (Figure 13e) and DOK4-PTK2 (Figure 13f). GRB7-OLIG1

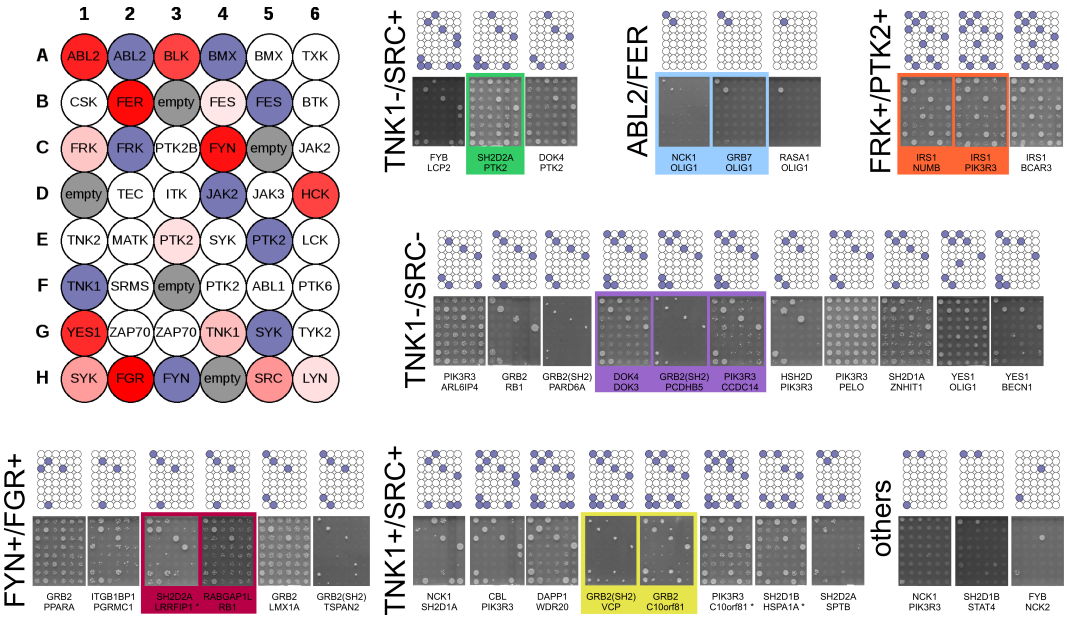


Figure 14: Phosphotyrosine-dependent interactions have different profiles in the kinase plate assay. Clusters of kinase plate assay results. For each interaction, pictures of the yeast colonies are shown, as well as a schematic representation of the spots considered for clustering. Clusters are labeled by the most distinguishing characteristics. Exemplars are colored for each cluster. Three interactions that did not cluster are shown in the lower right corner. In the top left corner, there is a schematic showing the arrangement of kinases in the kinase plate assay. For active kinases, the degree of redness indicates the activity (i.e. the number of interactions growing with the kinase), kinase-deficient mutants are colored blue and empty vector controls are grey. Note that for the interactions SH2D2A-LRRFIP1, PIK3R3-C10orf81 and SH2D1B-HSPA1A (marked by an astrisk(*)), there are colonies that were not considered for clustering.

grows with ABL2 and FER only. DOK4-PTK2 grows with six kinases, but not with ABL2 or FER. To provide another example, HSH2D-PIK3R3 (Figure 13g) grows with BLK, FYN, YES1 and FGR, that also allow DOK4-PTK2 growth, as well as with FER. It does not grow with HCK or SRC, that allow DOK4-PTK2 growth, or with ABL2.

In total, 37 interactions have been successfully assayed with two different kinase vectors in two independent replicates each. The resulting patterns (Figure 14) were evaluated in terms of growth with the fourteen kinases that gave rise to most of the signals and clustered using affinity propagation clustering (Frey & Dueck 2007). Affinity propagation clustering is a method for the identification

of groups of similar elements in many data sets. These groups are called clusters. In an iterative process, so-called examplars are found. The examplars are the most representative members of each cluster. For each pair of interactions, the squared number of common kinases, divided by the product of the number of interactions of both interactions, was used as a proxy for similarity. The resulting clusters are shown in Figure 14, the exemplars are marked by colored boxes. There are three small, specific clusters of three interactions each, three larger and more general clusters and three interactions that did not cluster. The first small cluster (Figure 14, green) seems to be defined by growth with SRC, but not with TNK1 or ABL2. The interactions in the next cluster (Figure 14, blue) are very clearly set apart from the others, because they grow with FER and ABL2 only. The third cluster (Figure 14, orange), in contrast, is made up by three interactions with at least twelve kinases, the most distinguishing ones being FRK and PTK2. Notably, SYK and FES are not seen. The three larger clusters are less specific. The first one (Figure 14, red) can best be described by the absence of specific kinase signals. Its exemplars grow with FYN, FER, FGR and ABL2, and all members of the cluster grow with at least three of the four and not more than one other kinase. The last two clusters encompass the majority of the clustered interactions. The most notable difference between the two is the absence of SRC and TNK1 from the larger one (Figure 14, purple). All interactions in the smaller one (Figure 14, yellow) have at least one of them. Interestingly, almost all interactions growing with FRK are in the "IRS1-NUMB/IRS1-PIK3R3" (orange) or in the "GRB2-VCP/GRB2-C10orf81" (yellow) cluster. Some of the interacting proteins demonstrate fidelity towards the clusters, but fail to fully explain the clusters. The three small clusters reflect to a high degree interactions with one particular protein. The first one (Figure 14, green) has both PTK2 interactions. The second one (Figure 14, blue) consists of three of the four OLIG1 interactions, the third one (Figure 14, orange) of all three IRS1 interactions. Apart from that, the two YES1 interactions can be found in the purple cluster and the two C10orf81 interactions are in the yellow cluster. The two proteins occurring most in the data set, GRB2 and PIK3R3, do not appear to be very cluster-specific, even though almost all PIK3R3 interactions grow with ABL2, BLK, FER, FYN, HCK, YES1 and FGR. In conclusion, the analyzed interactions showed principally similar but

appreciably different kinase patterns. The fact that they are similar indicates different levels of kinase activity. The fact that they are appreciably different as well indicates specificity of protein interactions towards kinases. The specificity is only partly explained by just one of the interacting proteins.

3.3 Bait and prey genes share process and function annotations

Next, we wanted to test our data in a biologically meaningful context. To do this, we tested how well our data agree with previous annotations in terms of biological role, localization and enzymatic activity. Towards this end, we used the Gene Ontology (GO) term annotations. Gene Ontology is a project that aims to characterize every gene using a clearly defined vocabulary (Ashburner et al. 2000). It uses a directed acyclic graph of terms linked by relationships, like ”is-a”. The ontology has three parts rooted in the most general terms, ”Cellular Component”, ”Molecular Function” and ”Biological Process”. Cellular Component collects terms related to subcellular localization, like nucleus or even specific complexes. Molecular Function describes low-level activities of genes and gene products, like enzymatic activities or specific binding abilities. Biological Process terms deal with more abstract roles the genes play on a systemic level, like eye development or apoptosis.
In principle, each gene is associated each appropriate GO term. Statistically, pairs of genes that interact in any way, especially genes whose products interact physically, are expected to have more similar GO terms than randomly paired genes. While this analysis has been conducted successfully in protein interaction research (Lehner & Fraser 2004), it produces statistically meaningful results only with very large numbers of interactions. There are several reasons for this. The first one is that some genes are more intensively researched than others influencing the degree of specificity of the annotated GO terms. Another one is that some GO terms are very common, potentially masking significant signals from less common GO terms. To address these issues instead of simply counting GO term matches, semantic similarities were calculated following the method of Resnik (Resnik 1995) implemented in the GOSemSim package for R (Yu

et al. 2010). To avoid creating bias in the selection of random interactions, the network of phosphorylation-dependent interactions and the complete network have been randomized in a way that preserves each proteins degree (Giot et al. 2003). If the average semantic similarity of interacting pairs of genes in the real networks is substantially different from the distribution of the average semantic similarity of interacting pairs of genes in the randomized networks, then we conclude that the network comprises mostly true interactions. The opposite is not true, neither can the method be applied to single interactions.
Figure 15 shows the average semantic similarity distribution for the randomized network with and without phosphorylation-independent interactions in blue and purple, respectively. The average semantic similarity of the original networks is indicated by arrows of the same color. For comparison, two large human yeast two-hybrid data sets, published by Rual et al. (2005) and Stelzl et al. (2005), were analyzed the same way and plotted in yellow and dark green, respectively. For all three ontologies, the average semantic similarities of both networks, with

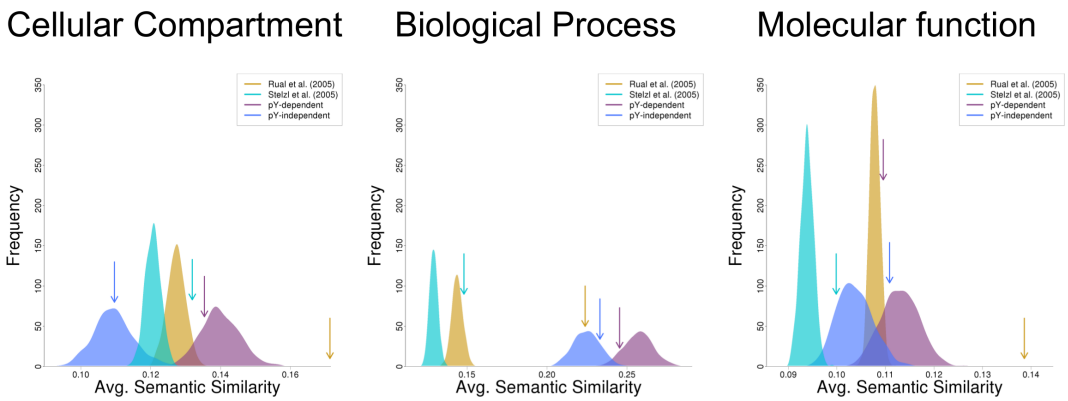


Figure 15: There is no significant enrichment in semantic similarity of Gene Ontology terms among interactions in the phosphotyrosine-dependent protein-protein interaction network. Interaction semantic similarity plots for the three Gene Ontologies "Cellular Compartment", "Biological Process" and "Molecular Function". One thousand randomized instances of the network of phosphotyrosine-dependent protein-protein interactions (purple) have been generated and the semantic similarity among interactions determined. The frequency is plotted against the average semantic similarity. The average semantic similarity of the original network is indicated by an arrow. For comparison, the values for the network of independent interactions and two high throughput yeast two-hybrid data sets presented by Stelzl et al. (2005) and Rual et al. (2005) are shown in blue, green and yellow, respectively.

and without independent interactions, clearly lies within the distribution of the randomized networks, while both reference networks show a clear difference between real and randomized networks, as has been reported before (Venkatesan et al. 2009). Interestingly, for two of the three ontologies, Biological Process and Molecular Function, the average semantic similarities for the randomized networks are much higher than the values for the reference sets. This might indicate that the prey set is too homogeneous to result a strong decrease in average semantic similarity through randomization in comparison to the real network. This would in turn mean that certain GO terms are over-represented in the set of interacting proteins as a whole.

Therefore we next analyzed the set of identified prey proteins for functional over-representation. Since the bait set is biased, because the bait genes have been selected according to a specific molecular function, that is known to be important in specific biological processes, like regulation of growth and development, only the prey set was analyzed for GO term over-representation. Because the different levels of the Gene Ontology contain largely overlapping information, we used only the level 4 terms. We counted the number of genes for each GO term in the prey set and in the background set. We used the genes in the prey matrix as the background set and calculated the specific over-representation using the hypergeometric test. The over-representation is more significant the more prey genes share a specific annotation and the smaller the total number of genes share this annotation. For example, there are 35 prey gene annotated as "kinase binding", and only 12 genes annotated as "protein phosphorylated amino acid binding". Nevertheless, the latter is more significant, because there are only 30 genes in the prey matrix with this annotation and 328 genes annotated as "kinase binding" (Figure 16). The p-values were corrected for multiple hypothesis testing by multiplying each p-value with the number of tests (Benjamini & Hochberg 1995). The same method was used to calculate over-representation p-values for the bait set to confirm and measure the bias in the bait selection. The terms that are also significantly over-represented for the bait set are indicated by circles next to the bars. Larger circles indicate higher levels of significance.

Significant over-representation of GO terms is indeed observed for all three ontologies. For Biological Process (Figure 17), the most significantly over-repre-

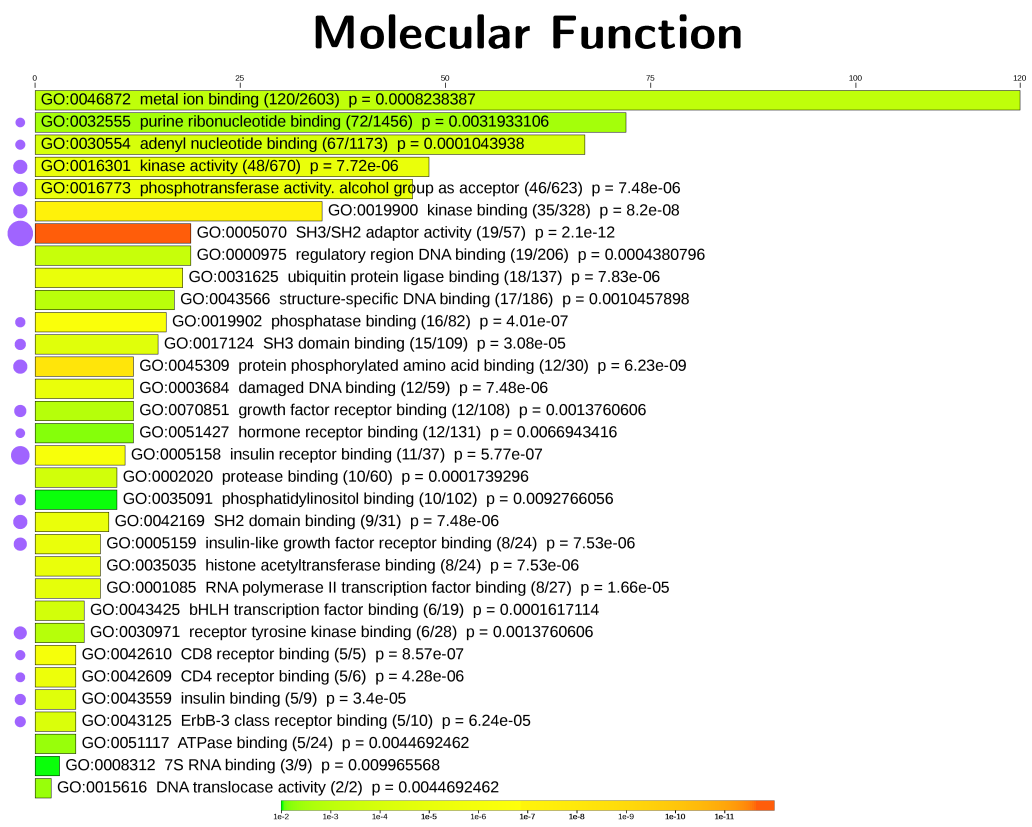


Figure 16: The interacting proteins are tied functionally to phosphotyrosine signaling. Over-representation of Gene Ontology Molecular Function terms in the set of prey proteins. The most over-represented Gene Ontology "Molecular Function" terms are shown with the number of genes with the annotation in the prey set, the number of genes with the annotation in the prey matrix, and the enrichment p-value calculated based on these values using the hypergeometric test. The length of the bars indicates the number of prey genes in the interaction network, the color indicates the over-representation p-value. Purple circles next to the bars indicate terms with significant over-representation in the bait gene set. Larger circles indicate higher levels of significance.

sented GO term has a p-value of 4×10^{-15}, for Molecular Function (Figure 16), it is 5.6×10^{-10}, and for Cellular Component (Figure 18) it is 2.2×10^{-5}. The number of over-represented GO terms common to the prey set and the bait set (marked by circles) is high for the Biological Process and Molecular Function ontologies and substantially lower for Cellular Compartment, as was suggested by the semantic similarity analysis of interacting pairs (Figure 15). In fact, the GO terms most highly over-represented and shared by the highest numbers of prey genes are "nucleus" and "nuclear lumen". This explains the fact, that

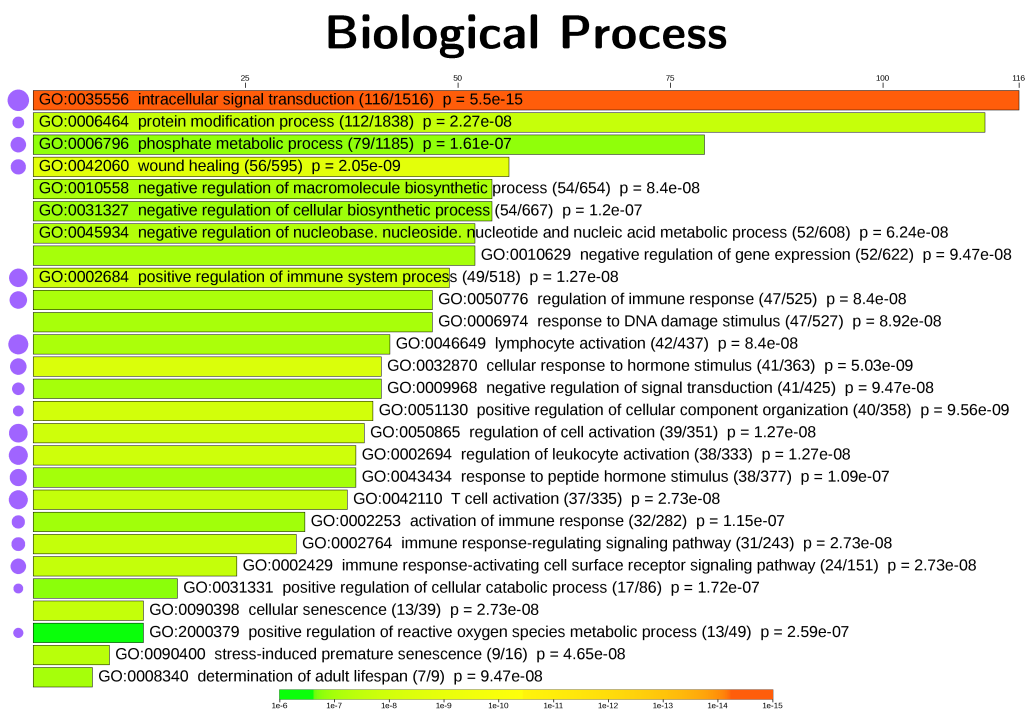


Figure 17: The interacting proteins are tied to phosphotyrosine-related processes. Over-representation of Gene Ontology Biological Process terms in the set of prey proteins. The most over-represented Gene Ontology "Biological Process" terms are shown with the number of genes with the annotation in the prey set, the number of genes with the annotation in the prey matrix, and the enrichment p-value calculated based on these values using the hypergeometric test. The length of the bars indicates the number of prey genes in the interaction network, the color indicates the over-representation p-value. Purple circles next to the bars indicate terms with significant over-representation in the bait gene set. Larger circles indicate higher levels of significance.

the average semantic similarity values for Cellular Compartment are lower than the values for the reference networks. For Molecular Function, the most highly over-represented GO terms are "SH3/SH2 adaptor activity", "protein phosphorylated amino acid binding" and "kinase binding". The term shared by most prey genes is "metal ion binding", but it is not the most significant one, because the number of metal ion-binding proteins in the background set is very high. It stemmed mainly from DNA binding proteins and kinases, which is another over-represented function. The over-representation of "SH3 domain binding" proteins is easily explained by the many proteins that contain both SH2 and SH3 domains. For Biological Process, the most over-represented and

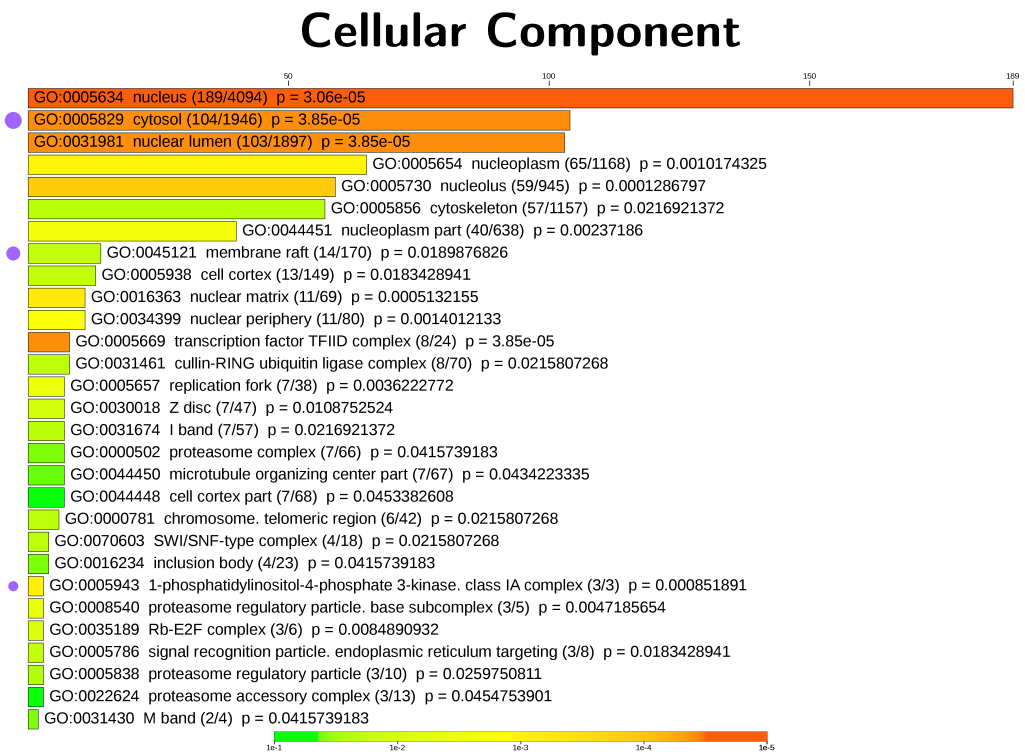


Figure 18: Most interacting proteins are cytoplasmic or nuclear. Over-representation of Gene Ontology Cellular Compartment terms in the set of prey proteins. The most over-represented Gene Ontology "Cellular Compartment" terms are shown with the number of genes with the annotation in the prey set, the number of genes with the annotation in the prey matrix, and the enrichment p-value calculated based on these values using the hypergeometric test. The length of the bars indicates the number of prey genes in the interaction network, the color indicates the over-representation p-value. Purple circles next to the bars indicate terms with significant over-representation in the bait gene set. Larger circles indicate higher levels of significance.

most common term is "intracellular signal transduction", which is also the most highly over-represented one for the bait set. Actually, most terms are common to the bait and the prey set except for negative regulation of processes related to growth and development. This might be due to mechanisms of down-regulation by binding or over-representation of transcription factors in the prey set. This agrees with the over-representation in nuclear and DNA-binding genes.
In conclusion, in terms of subcellular localization, the nuclear genes are most strongly over-represented in the prey set, while in the bait set, membrane-associated proteins are most strongly over-represented. In terms of biological and functional context, the prey set annotations agree very well with the biological

functions of the bait genes.

3.4 Bait and prey genes tend to belong to the same biological pathways

The Gene Ontology term over-representation analysis shows that genes important for the same biological functions are over-represented in the bait and

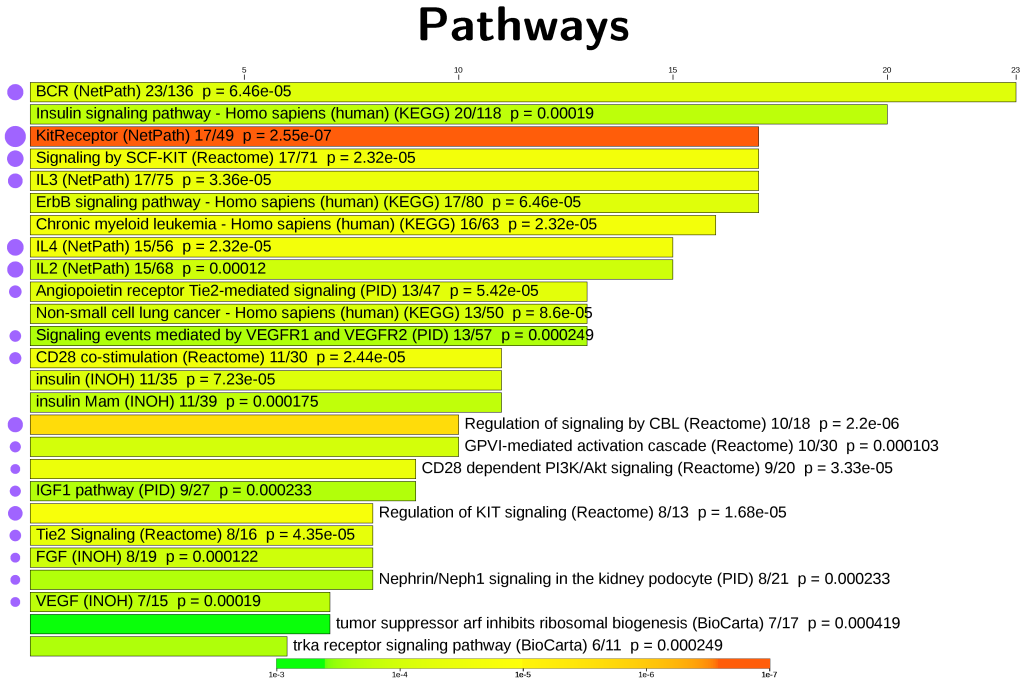


Figure 19: The interacting proteins are involved in phosphotyrosine signaling pathways. Over-representation of pathway annotations in the set of prey proteins. For the pathway annotation databases PID (Schaefer et al. 2009), Reactome (Matthews et al. 2009), NetPath (Kandasamy et al. 2010), Wikipathways (Kelder et al. 2012), KEGG (Kanehisa et al. 2012), INOH (Yamamoto et al. 2011) and Biocarta (`www.biocarta.com`) the most over-represented pathways are shown with the number of genes with the annotation in the prey set, the number of genes with the annotation in the prey matrix, and the over-representation p-value calculated based on these values using the hypergeometric test. The length of the bars indicates the number of prey genes in the interaction network, the color indicates the over-representation p-value. Purple circles next to the bars indicate pathways with significant over-representation in the bait gene set. Larger cirlces indicate higher levels of significance.

prey sets. This suggests that genes belonging to the same cellular pathways might also be over-represented. To determine pathway over-representation the ConsensusPathDB pathway meta-database (Kamburov et al. 2009) was queried for all genes overlapping between pathways in BioCarta (`www.biocarta.com`), INOH (Yamamoto et al. 2011), KEGG (Kanehisa et al. 2012), NetPath (Kandasamy et al. 2010), PID (Schaefer et al. 2009), Reactome (Matthews et al. 2009) or Wikipathways (Kelder et al. 2012) and the sets of bait, prey and background GeneIDs already used for the GO term over-representation analysis. Over-representation was calculated using the hypergeometric test, also as for the Gene Ontology term over-representation analysis.
Figure 19 shows the over-represented pathways among the prey genes. There is a large overlap between the over-represented pathways in bait genes and prey genes. The over-represented pathways are cytokine signaling, e.g. interleukin signaling, and growth factor receptor signaling pathways, like BCR, Kit Receptor, ErbB, VEGFR, Tie2, insulin and FGF signaling, as is expected for genes related to phosphotyrosine signaling (Lemmon & Schlessinger 2010). Interestingly, there are several insulin-related pathways significantly over-represented in the prey genes, but not the bait genes. Note that pathways with similar names defined by different sources are often quite different (Kamburov 2012). This is exemplified by Kit receptor signaling defined by NetPath and Wikipathways. Both are over-represented in the prey set, although the NetPath p-value is two orders of magnitude lower, but for the bait set the NetPath Kit receptor pathway is the most over-represented one, while the Wikipathways Kit pathway fails to reach significance. This is why it is important to query several pathway databases and judge the composite result (Kamburov et al. 2011). In conclusion, pathways related to phosphotyrosine signaling are over-represented in the prey set, as well as the bait set, demonstrating the relevance of the biological context of the presented data set.

3.5 Bait and prey genes are over-represented in the same interaction neighborhoods

Similar to the analyses of the genes interacting with phosphotyrosine-recognizing genes for over-representation of specific functions and pathways, we

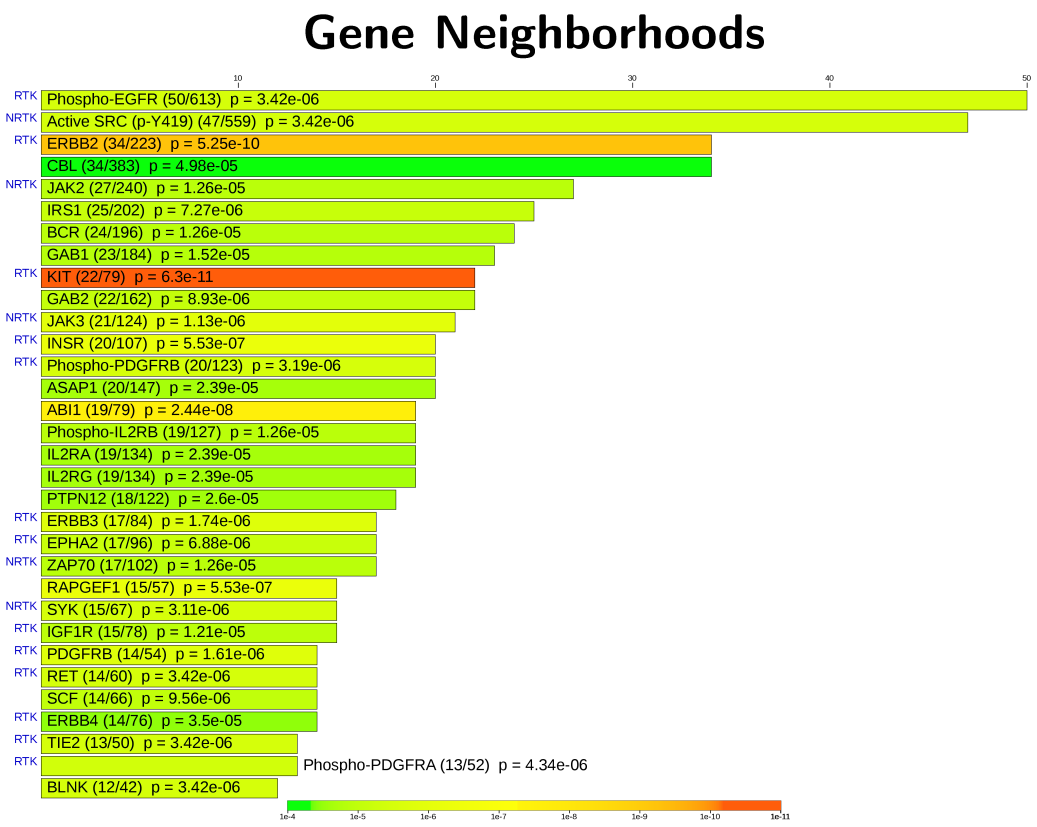


Figure 20: The interacting proteins are known to interact with phosphotyrosine signaling proteins, especially protein tyrosine kinases. Over-representation of network-based entity sets (NESTs) in the set of prey proteins. For all NESTs in the ConsensusPathDB database the most NESTs are shown with the number of genes with the annotation in the prey set, the number of genes with the annotation in the prey matrix, and the over-representation p-value calculated based on these values using the hypergeometric test. The length of the bars indicates the number of prey genes in the interaction network, the color indicates the over-representation p-value. NESTs centered around receptor tyrosine kinases (RTK) or non-receptor tyrosine kinases (NRTK) are indicated.

examined the network neighborhood of the proteins. We searched for over-represented genes that interact with specific other genes. This links the newly discovered interactions directly to existing protein-protein interaction knowledge (Kamburov et al. 2009) and thus provides valuable information about protein function. In the easiest case, the result should reflect the network hubs, but it can also show areas of high coherence not immediately apparent from a network determined with a specific experimental method. We observed the latter. The analysis showed many receptor tyrosine kinases, which are not easily

amiable to analysis by yeast two-hybrid analysis, because they are membrane proteins (Figure 20). The prey set was tested for over-representation of genes from radius one neighborhoods of all genes in the ConsensusPathDB protein interactions meta-database. The hypergeometric test with multiple testing correction was used to calculate the over-representation p-values. The most strongly over-represented neighborhood is the KIT receptor tyrosine kinase neighborhood, which is consistent with the result from the pathway analysis. Of the 31 over-represented neighborhoods, nine are receptor tyrosine kinases, another five are non-receptor tyrosine kinases, and well-known phosphotyrosine signaling genes. These findings provide strong support for the hypothesis that prey proteins interacting with phosphotyrosine-recognizing bait proteins are phosphorylated, based on the assumption that phosphorylation encompasses interaction with a kinase.

3.6 Cancer genes are strongly over-represented in the phosphotyrosine-dependent protein-protein interaction network

To test the applicability of the detected interactions in a biological or medical context, we assayed for over-representation of genes implicated in related human diseases. In the case of phosphotyrosine signaling, the most relevant disease association is cancer (Blume-Jensen & Hunter 2001). To assess the over-representation of cancer genes we refer to a cancer genes definition from a group of cancer specialists from the Wellcome Trust Sanger Institute (Futreal et al. 2004). The cancer gene list contains 487 genes found to carry mutations in cancer patients.

The analysis is presented in Table 7: The prey matrix has 13,931 non-autoactive prey strains. Among these are 303 cancer genes. There are 22 cancer genes among the 366 prey genes that have been found to interact with phosphotyrosine-recognizing domain-containing proteins. The chance of choosing 22 or more cancer genes from the prey matrix in 366 random draws without replacement are 2.4×10^{-5} (hypergeometric test). Considering the phosphorylation-dependent network alone, there are only 187 prey and 13 cancer genes, raising the odds

Table 7: Cancer genes are strongly over-represented among the protein-protein interactions. The cancer genes defined by Futreal et al. (2004) were tested for over-representation in the protein-protein interactions using the hypergeometric test. The analysis was performed on the basis of cancer genes found as prey and on the basis of interactions involving cancer genes as prey. For the first part, the number of total prey genes screened (prey matrix size), cancer genes screened (cancer prey genes total), total number of prey genes interacting with at least one bait (prey genes found) and the number of cancer prey genes interacting with at least one bait (cancer prey genes found) were determined for all interactions found in this study (complete network) and all phosphotyrosine-dependent interactions found in this study (phosphotyrosine-dependent only). Based on these values, the hypergeometric test was used to calculate the probability of seeing this kind of over-representation by chance alone (p-value). For the second part, the number of bait genes interacting with at least one prey gene (bait genes) was multiplied with the number of prey genes to calculate the total number of interactions (interactions possible) and the number of interactions involving cancer prey genes (cancer interactions possible) that could have been found. These values, and the actual number of total interactions (interactions found) and interactions involving cancer prey genes (cancer PPIs found) found in this study were used to calculate the probability of seeing this kind of enrichment by chance alone (p-value).

	complete network	phosphotyrosine-dependent only
prey matrix size	13931	13931
cancer prey genes total	303	303
prey genes found	366	187
cancer prey genes found	22	13
p-value	2.38×10^{-5}	2.83×10^{-4}
bait genes	70	59
interactions possible	975170	821929
cancer interactions possible	21210	17877
interactions found	628	292
cancer PPIs found	64	35
p-value	1.96×10^{-23}	1.13×10^{-15}

of a purely random occurrence to 2.8×10^{-4}. This probability is even lower when interactions are considered, not just whether a prey gene is present in

the set. In this case, the pool of possible interactions is the number bait genes multiplied by the number of prey genes in the matrix (13,931), the number of possible cancer gene interactions is given by the number of bait genes times the number of cancer genes in the matrix (303). In the complete set, there are 628 protein-protein interaction found with 70 bait genes. Sixty-four of these have cancer genes as prey. This corresponds to a probability of 2×10^{-23}. The 292 phosphorylation-dependent interactions, found with 59 bait genes, contain 35 interaction with cancer genes as prey. The probability of this happening randomly is 1.1×10^{-15}. The interaction-based analyses produce lower p-values not only because statistics on bigger numbers usually produce lower p-values, but also because the average number of bait geness that the cancer prey genes interact with is 3.27, which is significantly higher than the network average of 2.34 (Wilcoxon rank sum test with continuity correction, $p = 0.002774$).
In conclusion, there is a clear over-representation of genes implicated in cancer in the proteins interacting with phosphotyrosine-recognizing domain-containing proteins. The over-representation is more pronounced for the phosphorylation-dependent network than for the complete network. This suggests that the protein-protein interaction data set is biologically meaningful and may contain many relevant relationships that can be exploited for medical applications.

3.7 SH2 gene interactions can partially be accounted for by peptide binding motifs

We screened a comprehensive set of phosphotyrosine-recognizing domain-containing proteins for phosphotyrosine-dependent protein-protein interactions using full length human proteins. For many protein domains, especially SH2 domains, linear consensus peptide binding motifs based on peptide binding assays have been reported previously (Liu et al. 2006). Typically, in such an assay peptides of ten to fifteen amino acids, or mixtures of peptides with randomized amino acid positions, are assayed for the binding of purified domain constructs in vitro (Songyang et al. 1994, 1993; Lupher et al. 1997; Bunnell et al. 2000; Beebe et al. 2000; Huang et al. 2008). Since the proteins in this study, both the domain-containing and the recognized ones, were employed as full-length constructs, it is interesting to ask how well the results from the yeast

two-hybrid screen agree with the results of the peptide-based assays.

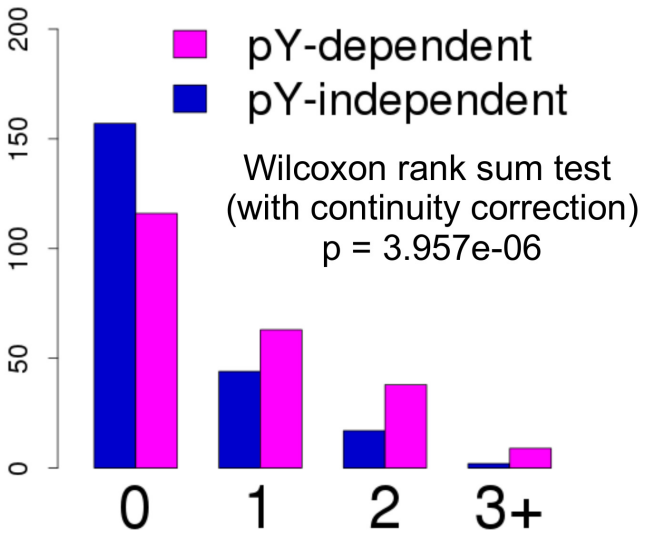
There are at least two possible approaches to this question. The first one is to search the amino acid sequences of the proteins in the prey matrix for the respective motifs and to assess the number of interactions with these genes as prey. This approach carries with it several problems. First, because of the high false negative rate in protein-protein interaction assays, distinguishing non-binding motifs from missed interactions is very hard, if not impossible. Second, tyrosine residues might be buried inside the protein, so the domain-containing protein cannot recognize them. Third, even if the motif is accessible, the respective tyrosine residue might not be phosphorylated in the assay system. The second one is to consider only interactors of each domain-containing protein and look for the presence of a known or novel motif. This approach is less influenced by false negative data and is thus more appropriate for yeast two-hybrid data (Weisberg 2010). The second and third problem still apply but are less problematic in this case. We simply assumed that all tyrosine residues are exposed and phosphorylated. If the putative phosphorylated protein does not contain a motive, then we can conclude that there must be a mechanisms of binding that is not governed by peptide motifs. The motifs collected by Liu et al. (2006) were used as literature consensus motifs. The complete set of phosphotyrosine-dependent interactions was analyzed rather than interactors of selected domain-containing proteins because of the large size of the data set, the natural heterogeneity in terms of numbers and size of interaction partners per domain-containing gene, and in terms of numbers of potential recognition sites per putative motif-containing protein as well as biases based on general sequence homology rather than motif-driven convergence.

By comparing the set of phosphorylation-dependent interactions to the complete

Figure 21: Phosphotyrosine-dependent interactors contain more known SH2 binding motifs than independent ones and are phosphorylated more often.

(Top) Number of positions matching known SH2 binding consensus motifs. For each protein interacting with an SH2 protein that has a consensus motif (defined by Liu et al. (2006)) the highest number of matching amino acid residues has been determined. The number of interaction partners is shown by best matching tyrosine motif for the phosphotyrosine-dependent and independent network.

Symbol	Motif	Number of matching positions: pY-dependent 1	pY-dependent 2	pY-dependent 3+	pY-independent 1	pY-independent 2	pY-independent 3	Reference
ABL2	[pY][E/T/M][N/E/D][P/V/L]	**2**	**1**					Songyang et al., 1993
CBL	[N][X][pY][S/T][X][X][P]	**3**	**5**		**4**	**1**		Lupher et al., 1997
CRK	[pY][D/K/N][H/F/R][P/V/L]	**16**	**19**	**7**	**18**	**10**		Songyang et al., 1993
FYN	[pY][E/T][E/D/Q][I/V/M]				**1**	**3**		Songyang et al., 1993
GRB10	[F/Y][pY][E/T/Y/S][N][I/L/V/P/T/Y/S]	**1**		**1**				Rodriguez et al., 2004
GRB2	[pY][I/V][N][I/L/V]	9	6		1	2	1	Rodriguez et al., 2004
	[pY][Q/Y/V][N][Y/Q/F]	7	4		1	1		Songyang et al., 1994
	[V][pY][Q][N][W/F]	6	2		1	1		Dente et al., 1997
		10	**8**		**1**	**3**	**1**	
GRB7	[F/Y][pY][E/T/Y/S][N][I/L/V/P/T/Y/S]		**2**	**1**				Rodriguez et al., 2004
LCK	[pY][E/T/Q][E/D][I/V/M]	**1**	**1**		**1**			Songyang et al., 1993
PIK3R3	[pY][M/I/V/E][X][M]	22	7		16	1		Songyang et al., 1993
	[pY][M/L/I][X][M]	22	2		17	2		Songyang et al., 1993
		24	**7**		**18**	**3**		
PTPN11	[I/L/V][I/L/V][I/F/V][pY][T/V][I/L][I/L/V/P]			**1**				Rodriguez et al., 2004
PTPN6	[L][Y/H][pY][M/F][X][F/M]	1						Beebe et al., 2000
	[pY][F][X][F/P/L/Y]	1						Songyang et al., 1994
	[V/I/L][X][pY][A][X][L/V]	1	1					Beebe et al., 2000
		1	**1**					
SH2D1A	[T][I][pY][X][X][V/I]	**2**	**1**		**4**			Poy et al., 1999
SH2D1B	[T][I][pY][X][X][V/I]	**1**			**2**	**1**		Poy et al., 1999
SH3BP2	[pY][E/M/V][N/V/I][X]	**2**						Songyang et al., 1994
SRC	[pY][EDT][ENY][IML]		**1**			**1**		Songyang et al., 1993
STAT3	[pY][X][X][Q]	**15**			**5**			Stahl et al., 1995
SYK	[pY][Q/T/E][E/Q][L/I]		**1**					Songyang et al., 1994
TNS1	[pY][D/E][N][I/F/V]						1	Auger et al., 1996
	[pY][E][N][F/I/V]						1	Songyang and Cantley, 1995
							1	



Ratio of known phosphosites

# of matches	pY-dependent	pY-independent
3+	40%	0%
2	33%	24%
1	18%	15%
0	23%	17%

(Bottom left) Histogram of the data presented in the top table. The distributions are significantly different with an error probability of $p = 3.769 \times 10^{-7}$ (as determined by a Wilcoxon rank sum test with continuity correction).
(Bottom right) Percentages of best matching sites known to be phosphorylated. For each best matching site, the RefSeq, UniProt and PhosphoSite databases have been searched for reports of phosphorylation in human or mouse.

set, statistical differences were found and quantified. Figure 21 (top) summarizes all literature motifs that have been found in the interaction set. For each domain-containing gene and each interactor, all tyrosine motifs, i.e. all tyrosine residues with surrounding residues, were compared to the consensus motif(s). The amino acids that were common to the consensus motif were counted and the highest count was defined as representing the most probable binding site. The purple and blue columns in Figure 21 (top) represent the counts of matching residues for phosphorylation-dependent and independent interactions, respectively. Because there is no obvious extrinsic reference and because of the high density of the phosphorylation-dependent interaction set, the set of independent interactions was used as a negative reference. Figure 21 (bottom left) shows the distribution of consensus matching-amino acid residues for all pairs of domain-containing proteins and interaction partners in the phosphorylation-dependent and independent set in purple and blue, respectively. The distribution corresponding to the phosphorylation-dependent interaction set is clearly shifted to the right. Since the values are not distributed normally, the Wilcoxon rank sum test was used to evaluate the significance of this difference. According to this test, the probability that both distributions are samples drawn from the same underlying distribution is $p = 3.769 \times 10^{-7}$. This analysis suggests that the p-dependent interaction partners contain more linear recognition motifs that may serve as recognition sites for their SH2 domain-containing partners. We also asked whether we can find known phosphorylated tyrosine residues in the putative recognition sites. The expectation is that tyrosine residues that are recognized by SH2 domains are phosphorylated in some type of human cell under certain conditions. This means, that known phosphorylation sites should be highly over-represented in the set of recognized tyrosine residues. Since the specific interaction sites are not known, the motifs matching the SH2 domain consensus binding motifs were used as proxies (Figure 21, bottom right). The known phosphorylation sites collected in the PhosphoSite, RefSeq and SwissProt databases were extracted for all prey proteins and their mouse homologs, assuming that tyrosine phosphorylation is conserved well between mouse and human. For each interaction, all tyrosine residues in the putative phosphorylated protein were categorized according to the number of surrounding residues matching the respective consensus motif. For phosphotyrosine-dependent inter-

actions, the percentage of known phosphorylation sites among tyrosine motifs matching the consensus motif in at least two positions is substantially elevated. The background level for non-matching motifs is around 18%-23%. For motifs matching in two or more than two positions the percentages are 33% and 40%, respectively. It must be noted that a similar trend can be observed for phosphorylation-independent interactions. The background level of 15%-17% is substantially lower than for phosphorylation-dependent interactions. For motifs matching in two positions, it is elevated to 24%. For three matching positions, the number of cases is very low and none of the two cases are reported to be phosphorylated.

We wondered whether the effect we observed is simply due to a difference in amino acid composition of the proteins. Therefore, we next considered the numbers of total and phosphorylated tyrosine residues in the two data sets. The average number of phosphorylated tyrosine residues reported is 2.17 per protein for the putative phosphorylated proteins in the phosphotyrosine-dependent interaction set and 1.25 for the proteins interacting with domain-containing proteins in the independent set, while the average total number of tyrosine residues is similar for both sets (phosphorylation-dependent: 12.5; independent: 10.3). The average percentage of tyrosine residues reported to be phosphorylated per protein is 14.4% and 10.4% for the phosphorylation-dependent and the independent set, respectively. Interestingly, these numbers differ from the quotients of phosphorylated tyrosine residues and the total number of tyrosine residues in the complete sets, which are 17.4% and 12.2% for the phosphorylation-dependent and the complete set, respectively. This suggests that the ratio of phosphorylated tyrosine residues is higher on average for proteins containing fewer tyrosine residues and that tyrosine phosphorylation is more protein-specific than site-specific, i.e. if a tyrosine phosphorylation site is observed for a protein, it is more likely that another tyrosine site in the same protein is observed to be phosphorylated. Thus, while proteins in the dependent and independent prey set differ in regard to tyrosine phosphorylation, there is no reason to believe that this difference may have caused the observed effects. The fact that the background levels of tyrosine phosphorylation are elevated in the interaction-centered analysis indicates that the proteins that are known to be phosphorylated have more interactions on average.

In conclusion, new SH2 domain binding peptide motifs could not be found, however, the SH2 domain consensus binding motifs reported in the literature are enriched in the phosphotyrosine-dependent interaction set and coincide with residues known to be phosphorylated in vivo. On the other hand, linear peptide motifs are not prominent in our data and are not sufficient to explain all of the phosphotyrosine-dependent interactions.

3.8 Systematic validation of phosphotyrosine-dependent protein interactions in mammalian cell culture suggests very high quality

The quality of protein-protein interaction data sets can be tested experimentally by assaying individual interactions with an independent protein interaction detection method, ideally one as different from the original method as possible (Braun et al. 2009). For yeast two-hybrid data, co-immunoprecipitation is an appropriate method, because it works in human cells, as opposed to yeast cells and is a complex-based method instead of a reporter activity-based, transient one (Phizicky et al. 2003). To validate the protein interaction set detected in this study, a set of 169 interactions was assayed in a high throughput luminescence based co-immunoprecipitation assay, 147 produced usable results, i.e. both proteins were successfully expressed at levels that allow detection by western blotting. The use of this type of assay for the validation of yeast two-hybrid results is well-established (Horn et al. 2006; Stelzl & Wanker 2006). In this co-immunoprecipitation assay, one of the two proteins was Protein A-tagged, while the other one was fused to a firefly luciferase gene (Figure 21a). Both were transiently expressed in mammalian HEK293 cells. The cells were lysed and the Protein A-tagged proteins were allowed to bind to immunoglobulin G-coated microplates. After washing, the amount of the second protein bound is measured in a luminescence assay and compared to a non-binding control. Pairs for which the signal exceeded the background by a factor of at least two and by at least two standard deviations from triplicate values, were considered positive. A total of 76 interactions were successfully validated this way (Figure 21b). The strength of the resulting signals ranges from two-fold (which is the cut-off), like

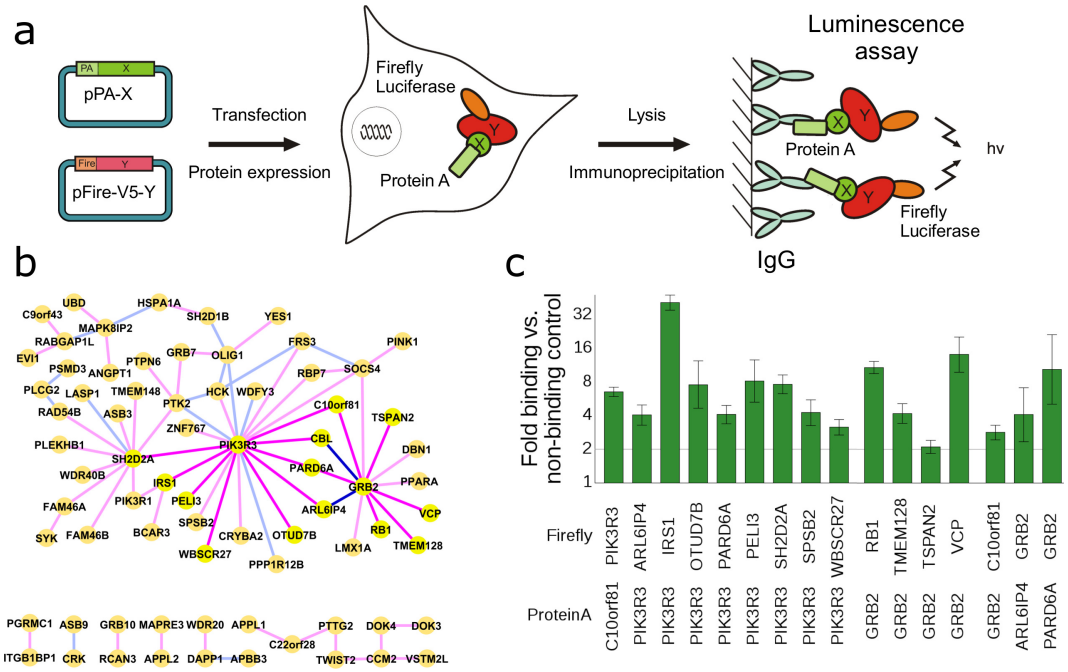


Figure 21: Seventy-eight protein-protein interactions were successfully validated in mammalian cell culture. (A) Principle of the Luminescence assay. The putatively interacting protein pair were genetically fused to a Protein A tag or a Firefly Luciferase gene and expressed in HEK293 from transiently expressed plasmids. The cells were lysed and the Protein A-tagged fusion proteins in the lysate were bound to immobilized IgG proteins. After washing, luciferase activity was determined. If the luciferase fusion protein was bound to the Protein A-tagged protein, luciferase activity could be detected.
(B) Network view of all interactions validated. Proteins are represented by nodes, links represent protein-protein interactions. Phosphotyrosine-independent interactions linking the nodes are shown in blue, phosphotyrosine-dependent interactions are purple. The emphasized links are shown in (C).
(C) Luminescence assay results for the interactions highlighted in (B). The fold binding over non-binding control is plotted. Error bars show standard deviations. Interactions were scored as positive if the binding signal was at least two-fold over non-binding control and differed by at least two standard deviations.

for PIK3R3-CBL, to more than thirty-fold binding for the PIK3R3-IRS1 protein pair (Figure 21c). Since protein interaction detection systems are orthogonal, which means that they each detect their own subset of all true interactions (Braun et al. 2009; Venkatesan et al. 2009), evaluation of data sets means comparing validation rates to validation rates obtained with a positive reference set rather than discarding interactions that fail to come back in the validation

assay as false. Interestingly, the validation rate of about 52% (Table 8), which is similar for phosphotyrosine-dependent and independent interactions, compares well to validation rates reported for this type of assay (Braun et al. 2009).
In conclusion, a substantial subset of the detected interactions has been assayed in mammalian co-immunoprecipitation experiments with validation rates that indicate high data quality.

Table 8: Validation rate in co-immunoprecipitation assays confirm the interactions high quality. One hundred and sixty-nine interactions were selected for co-immunoprecipitation from mammalian cell culture. Of these, 147 produced detectable levels of both interacting proteins. Interactions were considered validated if the luciferase signal exceeded non-binding control by a factor of at least two and by at least two standard deviations, assuming a multiplicative error model. The position of each ORF ("Protein A" or "Luciferase" for pPAReni-DM or pFireV5-DM, respectively), whether the interaction was phosphotyrosine-dependent in the yeast two-hybrid assay ("pY-dependent"), whether it could be validated ("validated"), the binding signal strength ("Fold Binding") and its statistical reliability in terms of standard deviations ("Z-Score") are provided for each evaluable pair of ORFs.

Protein A — Luciferase	pY-dependent	validated	Fold Binding	Z-Score
IRS1 — PIK3R3	yes	yes	$186.12^{\times}\div 1.89$	8.23
PPP1R12B — PIK3R3	no	yes	$157.64^{\times}\div 1.34$	17.48
IRS1 — PIK3R1	yes	yes	$91.85^{\times}\div 1.43$	12.68
TSPAN2 — GRB2	yes	yes	$39.66^{\times}\div 1.98$	5.4
PARD6A — PIK3R3	yes	yes	$27.33^{\times}\div 1.41$	9.58
ZNF767 — PIK3R3	yes	yes	$20.27^{\times}\div 1.12$	26.55
PLEKHB1 — SH2D2A	yes	yes	$17.12^{\times}\div 1.12$	26.02
PGRMC1 — ITGB1BP1	yes	yes	$15.3^{\times}\div 1.46$	7.22
GRB2 — VCP	yes	yes	$14.15^{\times}\div 1.44$	7.26
MAPRE3 — APPL2	yes	yes	$13.49^{\times}\div 1.08$	35.69
CRYBA2 — PIK3R3	yes	yes	$13.25^{\times}\div 1.17$	16.7
SH2D2A — FAM46A	yes	yes	$12.37^{\times}\div 1.12$	22.66
FRS3 — SOCS4	no	yes	$12.13^{\times}\div 1.39$	7.54
PELI3 — PIK3R3	yes	yes	$11.85^{\times}\div 1.1$	26.69
MAPK8IP2 — UBD	yes	yes	$11.09^{\times}\div 1.33$	8.48

Table 8: Validation rate in co-immunoprecipitation assays confirm the interactions high quality. (continued)

Protein A — Luciferase	pY-dependent	validated	Fold Binding	Z-Score
FAM46B — SH2D2A	yes	yes	$10.94^{\times}\div 1.09$	27.55
PARD6A — GRB2	yes	yes	$10.38^{\times}\div 2.05$	3.25
MAPK8IP2 — HSPA1A	no	yes	$10.22^{\times}\div 1.12$	20.27
PTK2 — HCK	no	yes	$9.83^{\times}\div 1.2$	12.68
FRS3 — PIK3R3	yes	yes	$8.57^{\times}\div 1.25$	9.55
PTK2 — PIK3R3	no	yes	$8.27^{\times}\div 1.21$	11.09
SH2D2A — PIK3R3	yes	yes	$7.95^{\times}\div 2.06$	2.87
PIK3R3 — SOCS4	yes	yes	$7.87^{\times}\div 1.26$	9.04
PIK3R3 — OTUD7B	yes	yes	$7.59^{\times}\div 1.64$	4.12
RABGAP1L — EVI1	yes	yes	$7.19^{\times}\div 1.15$	14.02
GRB2 — CBL	no	yes	$6.94^{\times}\div 1.1$	20.14
FRS3 — HCK	no	yes	$6.87^{\times}\div 1.2$	10.54
C10orf81 — PIK3R3	yes	yes	$6.56^{\times}\div 1.1$	19.75
FRS3 — FRS3	no	yes	$6.02^{\times}\div 1.14$	13.73
RAD54B — SH2D2A	yes	yes	$5.88^{\times}\div 1.15$	12.34
OLIG1 — PIK3R3	no	yes	$5.19^{\times}\div 1.23$	7.81
GRB7 — PTK2	yes	yes	$4.78^{\times}\div 1.69$	2.97
PTK2 — SH2D2A	yes	yes	$4.61^{\times}\div 1.12$	13.38
GRB2 — DBN1	yes	yes	$4.36^{\times}\div 1.34$	4.98
SYK — FAM46A	yes	yes	$4.27^{\times}\div 1.33$	5.05
CCM2 — VSTM2L	yes	yes	$4.24^{\times}\div 1.63$	2.96
DOK4 — DOK3	yes	yes	$4.22^{\times}\div 1.56$	3.22
GRB2 — TMEM128	yes	yes	$4.18^{\times}\div 1.22$	7.08
SPSB2 — PIK3R3	yes	yes	$4.11^{\times}\div 1.18$	8.55
ARL6IP4 — GRB2	no	yes	$4.08^{\times}\div 1.74$	2.53
PIK3R3 — ARL6IP4	yes	yes	$4.05^{\times}\div 1.23$	6.82
MAPK8IP2 — RABGAP1L	no	yes	$3.87^{\times}\div 1.27$	5.57
LASP1 — SH2D2A	no	yes	$3.87^{\times}\div 1.17$	8.52
HSPA1A — SH2D1B	yes	yes	$3.53^{\times}\div 1.24$	5.84

Table 8: Validation rate in co-immunoprecipitation assays confirm the interactions high quality. (continued)

Protein A — Luciferase	pY-dependent	validated	Fold Binding	Z-Score
PIK3R1 — SH2D2A	yes	yes	$3.37^{\times}_{\div}1.11$	11.29
TWIST2 — CCM2	yes	yes	$3.37^{\times}_{\div}1.21$	6.48
ASB9 — CRK	no	yes	$3.36^{\times}_{\div}1.67$	2.36
C22orf28 — APPL1	yes	yes	$3.35^{\times}_{\div}1.25$	5.48
RAD54B — PLCG2	no	yes	$3.35^{\times}_{\div}1.59$	2.59
WDR20 — DAPP1	yes	yes	$3.34^{\times}_{\div}1.31$	4.53
PTPN6 — PTK2	yes	yes	$3.31^{\times}_{\div}1.75$	2.14
GRB2 — RB1	yes	yes	$3.29^{\times}_{\div}1.27$	5.02
PIK3R3 — WBSCR27	yes	yes	$3.15^{\times}_{\div}1.17$	7.25
OLIG1 — SH2D1B	no	yes	$3.1^{\times}_{\div}1.32$	4.08
OLIG1 — HCK	no	yes	$3.1^{\times}_{\div}1.27$	4.73
OLIG1 — YES1	yes	yes	$3^{\times}_{\div}1.19$	6.44
BCAR3 — IRS1	yes	yes	$2.95^{\times}_{\div}1.4$	3.2
GRB2 — PPARA	yes	yes	$2.87^{\times}_{\div}1.12$	9.49
GRB2 — C10orf81	yes	yes	$2.83^{\times}_{\div}1.16$	6.95
LMX1A — GRB2	yes	yes	$2.82^{\times}_{\div}1.48$	2.63
MAPK8IP2 — ANGPT1	yes	yes	$2.66^{\times}_{\div}1.2$	5.32
RBP7 — PIK3R3	yes	yes	$2.66^{\times}_{\div}1.15$	7.2
APBB3 — DAPP1	no	yes	$2.65^{\times}_{\div}1.32$	3.55
GRB2 — SOCS4	yes	yes	$2.59^{\times}_{\div}1.19$	5.55
PIK3R3 — WDFY3	no	yes	$2.49^{\times}_{\div}1.19$	5.15
PINK1 — SOCS4	yes	yes	$2.46^{\times}_{\div}1.11$	8.71
CBL — PIK3R3	yes	yes	$2.41^{\times}_{\div}1.52$	2.11
PIK3R3 — HCK	yes	yes	$2.38^{\times}_{\div}1.38$	2.68
OLIG1 — GRB7	yes	yes	$2.32^{\times}_{\div}1.31$	3.1
WDR40B — SH2D2A	yes	yes	$2.29^{\times}_{\div}1.15$	5.76
CCM2 — DOK4	yes	yes	$2.27^{\times}_{\div}1.4$	2.44
RBP7 — SOCS4	yes	yes	$2.24^{\times}_{\div}1.22$	4.07
RCAN3 — GRB10	yes	yes	$2.21^{\times}_{\div}1.1$	8.69

Table 8: Validation rate in co-immunoprecipitation assays confirm the interactions high quality. (continued)

Protein A — Luciferase	pY-dependent	validated	Fold Binding	Z-Score
SH2D2A — ASB3	yes	yes	$2.09^{\times}\div 1.16$	4.83
C9orf43 — RABGAP1L	yes	yes	$2.03^{\times}\div 1.14$	5.39
PSMD3 — PLCG2	no	yes	$2.02^{\times}\div 1.19$	4.09
SH2D2A — TSC1	yes	no	$3.31^{\times}\div 1.92$	1.84
ASB9 — PLCG2	no	no	$2.17^{\times}\div 1.86$	1.25
PTK2 — SRC	no	no	$1.96^{\times}\div 1.35$	2.24
GRB2 — WBSCR27	no	no	$1.94^{\times}\div 1.22$	3.37
CRK — SH2D2A	yes	no	$1.94^{\times}\div 1.19$	3.84
PIK3R3 — PPARA	yes	no	$1.87^{\times}\div 1.38$	1.94
RABGAP1L — WDFY3	yes	no	$1.79^{\times}\div 1.39$	1.79
APBB3 — PDPK1	yes	no	$1.76^{\times}\div 1.76$	1.01
OLIG1 — STAT5A	yes	no	$1.64^{\times}\div 2.45$	0.55
OLIG1 — GRB2	no	no	$1.58^{\times}\div 1.69$	0.87
RALBP1 — PLCG2	no	no	$1.58^{\times}\div 1.19$	2.58
BECN1 — YES1	yes	no	$1.57^{\times}\div 1.36$	1.47
FYN — CBLB	no	no	$1.56^{\times}\div 1.3$	1.72
PIK3R3 — VCP	yes	no	$1.53^{\times}\div 1.55$	0.97
MAPK8IP2 — SH2D1B	no	no	$1.51^{\times}\div 1.26$	1.76
OLIG1 — PTPN6	yes	no	$1.5^{\times}\div 1.52$	0.98
PLAGL2 — DAPP1	no	no	$1.49^{\times}\div 1.47$	1.03
LRRFIP1 — SH2D2A	yes	no	$1.45^{\times}\div 1.12$	3.4
STAT4 — CRK	yes	no	$1.39^{\times}\div 1.75$	0.59
CRK — CBLB	yes	no	$1.39^{\times}\div 1.43$	0.92
TWIST2 — PLCG2	no	no	$1.38^{\times}\div 1.21$	1.65
ELK1 — PLCG2	yes	no	$1.37^{\times}\div 1.45$	0.85
SH2D2A — LNX1	yes	no	$1.34^{\times}\div 1.26$	1.27
PIK3R3 — PACRGL	no	no	$1.31^{\times}\div 1.22$	1.34
MAPK8IP2 — TMEM128	yes	no	$1.29^{\times}\div 1.12$	2.21
RABGAP1L — DBN1	yes	no	$1.27^{\times}\div 1.24$	1.09

Table 8: Validation rate in co-immunoprecipitation assays confirm the interactions high quality. (continued)

Protein A — Luciferase	pY-dependent	validated	Fold Binding	Z-Score
CSK — PDPK1	yes	no	$1.24 \stackrel{\times}{\div} 1.26$	0.95
PIK3R1 — CRK	yes	no	$1.17 \stackrel{\times}{\div} 1.47$	0.42
C22orf28 — CRK	yes	no	$1.16 \stackrel{\times}{\div} 1.67$	0.29
VASP — FRS3	yes	no	$1.15 \stackrel{\times}{\div} 1.43$	0.4
PIK3R3 — C1orf135	no	no	$1.13 \stackrel{\times}{\div} 1.09$	1.43
EPS8 — RCAN3	no	no	$1.13 \stackrel{\times}{\div} 1.42$	0.34
PTK2 — SH2D1B	no	no	$1.11 \stackrel{\times}{\div} 1.34$	0.36
JAK3 — LNX1	yes	no	$1.08 \stackrel{\times}{\div} 1.38$	0.24
HSH2D — PIK3R3	yes	no	$1.07 \stackrel{\times}{\div} 1.73$	0.13
PIK3R3 — SRC	yes	no	$1.07 \stackrel{\times}{\div} 1.44$	0.19
TWIST2 — CRK	no	no	$1.05 \stackrel{\times}{\div} 1.73$	0.09
MAPK8IP2 — LOC492311	yes	no	$1 \stackrel{\times}{\div} 1.13$	0.03
GRB2 — MYOZ1	yes	no	$1 \stackrel{\times}{\div} 1.14$	0.01
LIX1 — APPL2	yes	no	$1 \stackrel{\times}{\div} 1.31$	−0.02
STAT4 — SH2D1B	yes	no	$0.99 \stackrel{\times}{\div} 1.32$	−0.04
RABGAP1L — RB1	yes	no	$0.97 \stackrel{\times}{\div} 1.42$	−0.08
OLIG1 — NUMB	yes	no	$0.96 \stackrel{\times}{\div} 1.23$	−0.18
LGALS9C — APPL2	yes	no	$0.95 \stackrel{\times}{\div} 1.24$	−0.24
DOK5 — SCOC	yes	no	$0.93 \stackrel{\times}{\div} 2.33$	−0.08
RABGAP1L — TSPAN2	yes	no	$0.92 \stackrel{\times}{\div} 1.14$	−0.62
PTK2 — DOK4	yes	no	$0.9 \stackrel{\times}{\div} 1.35$	−0.34
FRS3 — NUMB	yes	no	$0.88 \stackrel{\times}{\div} 1.13$	−1.05
CBL — PIK3R2	yes	no	$0.84 \stackrel{\times}{\div} 1.45$	−0.47
ZHX3 — SH2D2A	yes	no	$0.83 \stackrel{\times}{\div} 1.27$	−0.76
IRS1 — NUMB	yes	no	$0.83 \stackrel{\times}{\div} 1.12$	−1.62
OLIG1 — DOK4	no	no	$0.8 \stackrel{\times}{\div} 1.3$	−0.82
C8orf33 — MAPK8IP2	yes	no	$0.79 \stackrel{\times}{\div} 1.94$	−0.35
FRS3 — MATK	yes	no	$0.79 \stackrel{\times}{\div} 1.25$	−1.06
ELK1 — CRK	yes	no	$0.78 \stackrel{\times}{\div} 1.75$	−0.45

Table 8: Validation rate in co-immunoprecipitation assays confirm the interactions high quality. (continued)

Protein A — Luciferase	pY-dependent	validated	Fold Binding	Z-Score
ARFGAP1 — RABGAP1L	yes	no	$0.77 \overset{\times}{\div} 1.15$	−1.92
GRB2 — WDFY3	yes	no	$0.74 \overset{\times}{\div} 1.45$	−0.8
KLRAQ1 — SH2D2A	yes	no	$0.73 \overset{\times}{\div} 1.14$	−2.52
GRB2 — PACRGL	yes	no	$0.72 \overset{\times}{\div} 1.15$	−2.43
CBL — STAT5A	yes	no	$0.69 \overset{\times}{\div} 2.11$	−0.5
APPL1 — SH2D2A	yes	no	$0.65 \overset{\times}{\div} 1.23$	−2.11
PIK3R3 — EVI1	yes	no	$0.62 \overset{\times}{\div} 1.15$	−3.53
MPP5 — SOCS4	yes	no	$0.6 \overset{\times}{\div} 1.42$	−1.47
STAT2 — DOK4	yes	no	$0.58 \overset{\times}{\div} 1.24$	−2.51
C22orf39 — MAPK8IP2	yes	no	$0.56 \overset{\times}{\div} 1.89$	−0.92
CRK — LASP1	yes	no	$0.51 \overset{\times}{\div} 1.35$	−2.25
PIK3R3 — ZNF281	yes	no	$0.49 \overset{\times}{\div} 1.26$	−3.08
APPL1 — PIK3R2	no	no	$0.41 \overset{\times}{\div} 1.15$	−6.45
CRK — PIK3R2	yes	no	$0.38 \overset{\times}{\div} 1.21$	−5.08
LDHAL6B — SOCS4	yes	no	$0.22 \overset{\times}{\div} 1.57$	−3.4
CRK — SEMA4D	yes	no	$0.13 \overset{\times}{\div} 3.47$	−1.63

3.9 Binding-impaired SH2 domains corroborate phosphotyrosine-dependency in co-immunoprecipitation experiments

We performed a genome-scale yeast two-hybrid screen for phosphotyrosine-dependent protein-protein interactions. Further, we showed that the interacting proteins are functionally related and have a high validation rate in co-immunoprecipitation experiments from human cells. While the functional over-representation analyses showed a clear connection to tyrosine phosphorylation, the co-immunoprecipitation experiments did not. Therefore, we propose three

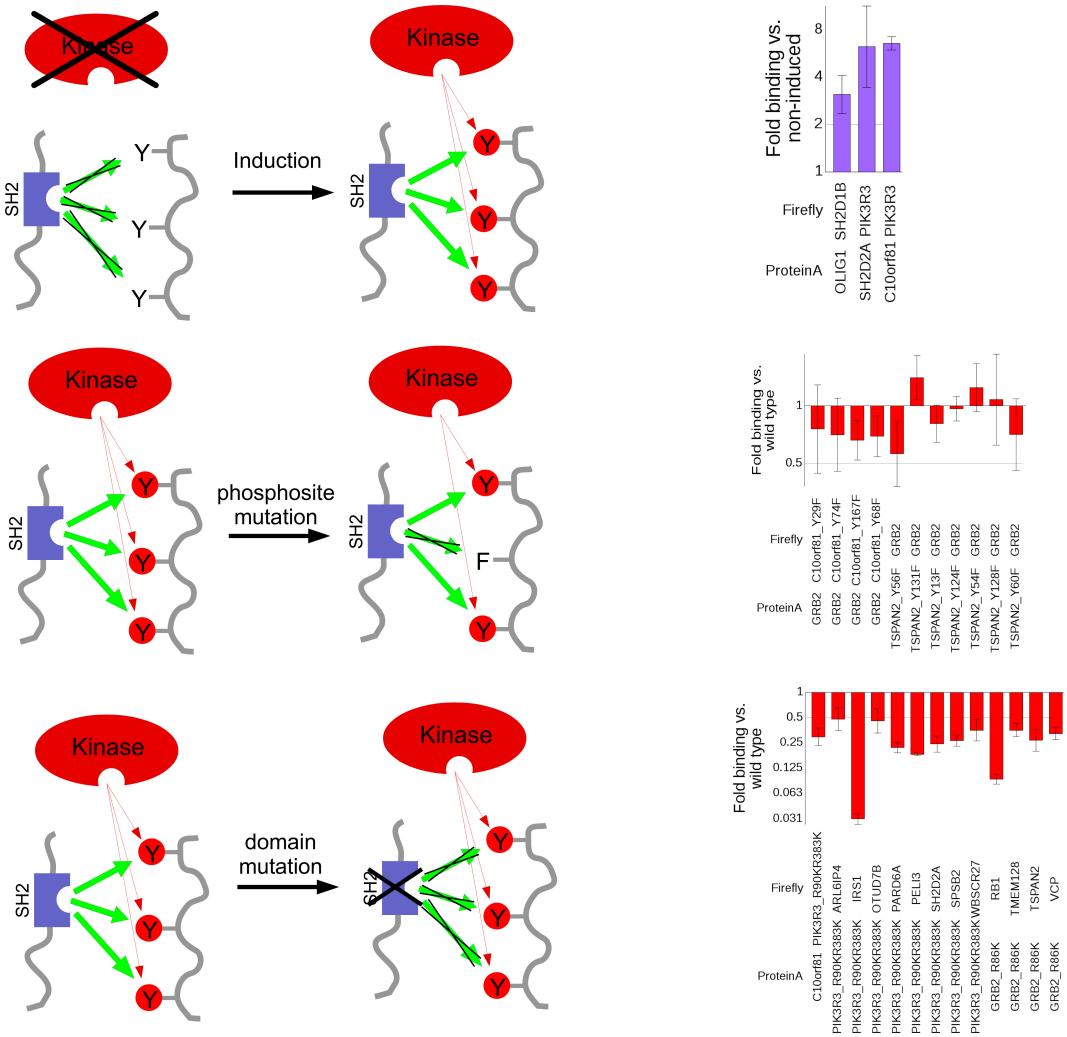


Figure 22: SH2 domain mutations prove phosphotyrosine-dependency in mammalian cell culture.
(Left) Schematic of three approaches to distinguish phosphotyrosine-dependent protein-protein interactions from independent ones. From top to bottom: Modulating tyrosine phosphorylation state by varying kinase expression; disruption of binding by replacing the target tyrosine residue by one that cannot be phoshporylated; mutation of the SH2 domain rendering it unable to bind to phosphotyrosine.
(Right) Corresponding experimental results.
(Top) All 75 interactions scoring positive in the luminescence assay were tested with and without induction of an ABL2 kinase. Plotted is the relative binding signal compared to the non-induced result. Three showed a significant and substantial increase when induced. (Middle) Single tyrosine-to-phenylalanine mutations have been generated for all tyrosine residues in two proteins interacting with GRB2 and tested for binding in the Luminescence assay. Plotted is the relative binding signal compared to the wild type result. Several mutations caused a decrease in binding signal but none of them led to a significant

different ways of demonstrating that protein-protein interactions are phosphorylation-dependent (Figure 22). The first one is comparing the binding signals of two interacting proteins with or without kinases or with different kinases or different kinase expression levels (Figure 22, top). In the yeast two-hybrid retest experiments, this principle was exploited comparing yeast growth between experiments using kinase vectors with and without active kinases.
In the co-immunoprecipitation assays, a large fraction of interactions from a comprehensive set of phosphotyrosine-dependent protein-protein interactions was validated successfully. All of these were tested in HEK293 cells carrying a stably integrated ABL2 kinase gene under the control of a tetracycline-dependent promoter with and without tetracycline induction. Only three of them showed a clear difference upon the induction (Figure 22, top right). The most likely explanation for this finding is the high endogenous kinase activity in fast-growing cells such as HEK293. Clearly, one of the other two principles needed to be applied.
A phosphotyrosine-dependent interaction can be disrupted by mutating either the tyrosine residue or the phosphotyrosine-recognizing domain. For a successful tyrosine residue mutation strategy, the interaction has to follow a simple mode of binding involving only a single tyrosine residue. Further, it requires knowledge about which residue is recognized by the domain-containing protein. If this information is unavailable, all tyrosine residues need to be examined in the protein-protein interaction perturbation tests. In this case it provides this knowledge, but requires a much greater effort. We created phenylalanine mutant clones for all tyrosine residues in C10orf81 and TSPAN2, four in C10ORF81 and seven in TSPAN2, and tested the binding to GRB2 in the co-immunoprecipitation assay in comparison to the wild type proteins. While there is a tendency for a reduction in binding, all effect sizes were smaller than two-fold and statistically not significant (Figure 22, center right).
Mutating the phosphotyrosine-recognizing domain, on the other hand, proofed

---

decrease.
(Bottom) Phosphotyrosine binding-disrupting SH2 domain mutations have been generated for GRB2 and PIK3R3 and tested in the Luminescence assay. Plotted is the relative binding signal compared to the wild type result.

more efficient. In SH2 domains, there is a conserved arginine residue that is known to reduce the ABL1 SH2 domains phosphotyrosine binding affinity by more than 100-fold when mutated to lysine (Mayer et al. 1992). (Marengere & Pawson 1992) used a similar mutation to distinguish binding of the N-terminal and C-terminal RASA1 SH2 domain binding to the EGF receptor. In this study, employing SH2 domain mutants required a relatively lower cloning effort, because there are only one or two SH2 domains per protein and many tyrosine residues in most putative tyrosine-phosphorylated proteins (i.e. their interaction partners). Additionally, it allowed testing of all interactions that have been detected with the respective protein and SH2 proteins have more phoshotyrosine-dependent interaction partners than the putative tyrosine-phosphorylated proteins. This point is especially important, because some interactions are enabled by several tyrosine residues, if they are phosphorylated. For example, Kouhara et al. (1997) showed that there are four binding sites, each sufficient for interaction between GRB2 and FRS2. A disadvantage of this approach is that it can be applied to SH2 domains much better than to PTB domains. In PTB domains, no generally conserved residue that is strictly required for phosphotyrosine binding is known. Another problem, that applies to both mutational approaches, is that there are many interactions among domain-containing proteins. In these cases it is unclear which of the two proteins is phosphorylated and which one is recognizing the phosphorylated tyrosine residue. Nevertheless, it is easier to test all domains that might be involved than all tyrosine residues. For these reasons thirty phosphotyrosine-dependent protein-protein interactions that tested positive in co-immmnoprecipitation validation assays were tested with mutated SH2 domains (Figure 22, bottom). Twenty of them showed a significant decrease of binding by at least 50% compared to wild type binding. The phosphotyrosine-recognizing domain-containing genes chosen to demonstrate phosphorylation-dependency in mammalian cell culture were PIK3R3, GRB2, PIK3R1 and SOCS4. PIK3R3 and GRB2 were selected because they have a larger number of phosphotyrosine-dependent interaction partners, PIK3R1 for its high sequence similarity to PIK3R3 and SOCS4 because it also interacts with PIK3R3 and GRB2.

Thus, for a large fraction of phosphotyrosine-dependent interactions, we confirmed the requirement of the phosphotyrosine-recognizing domain for binding.

3.10 The phosphotyrosine-dependent protein interactions found in this study are mostly novel

A comprehensive phosphotyrosine-dependent protein-protein interaction data set has been collected in a genome-scale yeast two-hybrid screen and shown to be of high quality in mammalian cell culture. However, the luminescence-based co-immunoprecipitation experiments can only assess the quality of the detected interactions, but it does not allow drawing conclusions about the data completeness. In more technical terms, the validation assays help evaluate the precision but not the sensitivity of the screen (Wilson 1927).

We wanted to estimate the number of true interactions missed to calculate the false negative rate in our assay. Because the complete set of true interactions is unknown, a literature-curated interaction set is commonly used as a proxy (Simonis et al. 2009; Braun et al. 2009, for example). For the purposes of this analysis all possible interactions between successful bait genes were used as a defined interaction space. The protein interaction meta-database Consensus-PathDB (Release 19) (Kamburov et al. 2009) was used to find all interactions in the bait-bait interaction space annotated in the databases BIND, Biogrid, DIP, HPRD, InnateDB, IntAct, MINT, MIPS, NetPath and Spike. Two hundred and thirty-nine annotated interactions were obtained and controlled by manually surveying the references until at least one was found that contains evidence for a physical interaction between the two genes in question. This included e.g. GST pull-down, yeast two-hybrid or co-immunoprecipitation. The literature-curated interaction set contains 147 interactions that could be validated this way. This data set was compared to the set of interactions detected in this study (Figure 22). In this study, 81 protein-protein interaction were found in the interaction space, 51 phoshotyrosine-dependent and 30 independent ones. Of these, 15 overlapped with the literature-curated interaction set (red in Figure 22). This suggests a false negative rate of almost 90%, which is common for protein-protein interaction sets based on a single detection method (Braun

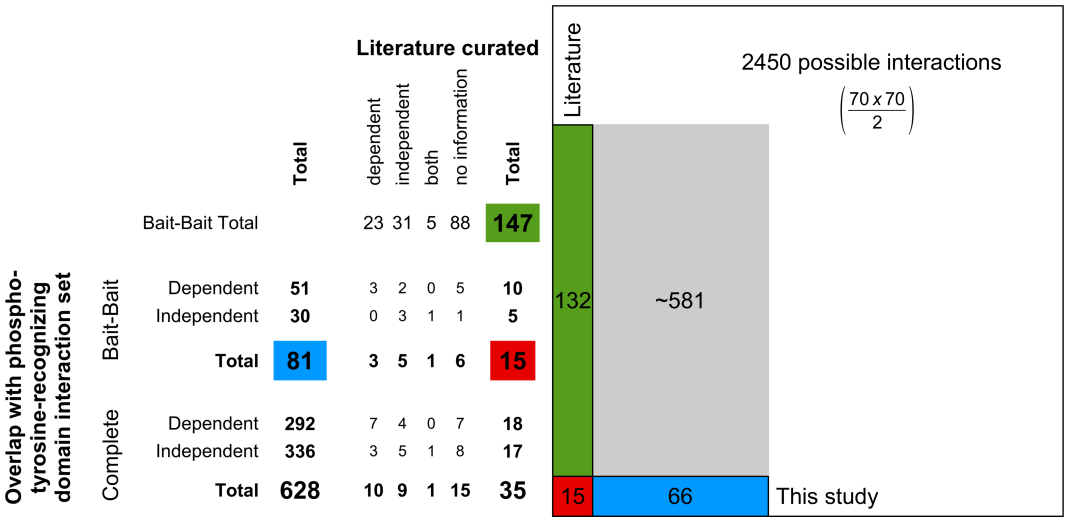


Figure 22: Most of the interactions found in this study are novel. (Left) Overview of the overlap of the interactions found in this study and interactions found in the literature. The ConsensusPathDB protein-protein interaction meta-database was queried for all possible interactions among successful bait proteins, as well as all other interactions found in this study. The retrieved articles were controlled and until an interaction was validated or until all article for that interaction were exhausted. Were applicable, information about phosphotyrosine-dependency was collected. Column denote literature results, rows results of this study. The fields are the respective overlaps. The colored fields contain data used on the right.
(Right) Graphical representation of the bait-bait interaction space, the space covered by literature interactions (green), the space covered by this study (blue), the overlap between the two (red) and the space of expected unknown interactions (grey), based on the assumption that the results found in the literature and the results of this study are independent and unbiased.

et al. 2009; Venkatesan et al. 2009).

The comparison to the literature allows the assessment of two other interesting parameters. The first one is the novelty of the interaction set (i.e. the degree to which the detected interactions expand the protein interaction knowledge), the other one is the number of interactions in the interaction space that have not been found yet (Figure 22, right). The novelty is approximated by the ratio of interactions in the screen not covered by the literature. Eighty-one interactions have been found in the bait-bait space. Fifteen of those overlap with the literature-curated data set. This means that only about 18.5% of the interactions are known, assuming there are no biases in the literature-curated set. The total number of true interactions in the bait-bait space can be estimated from the size

of the two interaction sets and the overlap. If we assume the interactions found in this study and the interactions in the literature-curated set to be two samples independently drawn from a common pool of true protein-protein interactions in the interaction space and the probability of finding an interaction to be uniform, the total number of interactions in the interaction space is given by the formula . This estimates the number of unknown interactions to be about five hundred and eighty-one (grey box in Figure 22).

To get a better estimate of the novelty, the same databases used to generate the literature-curated interaction set were queried for the rest of the detected interactions and controlled in the same way. Interestingly, there are only 35 interactions in the literature for the complete set of 628 interactions, which is only 5.6% making it considerably lower than the 18.5% in the bait-bait interaction set. This indicates a relatively strong research bias towards interactions among phosphotyrosine-recognizing domain-containing proteins.

For the yeast two-hybrid-derived data set, each interaction is detected as either phosphotyrosine-dependent or independent. For the literature-curated set, this information is not always available. For the purposes of this comparison, phosphorylation dependency was inferred for 20 of the 35 interactions, wherever reasonable. For example, if an SH3 domain alone is sufficient for a successful GST-pulldown experiment, it is reasonable to assume that the interaction is independent (Chan et al. 2003). Similarly, if an interaction can be disrupted replacing a tyrosine by a phenylalanine residue, it is probably phosphorylation-dependent (Elly et al. 1999). Interestingly, for some interactions, information pointing in both directions is available (Bandyopadhyay et al. 2010; Blagoev et al. 2003; Howlett et al. 1999; Elly et al. 1999; Kotani et al. 1998; Scholz et al. 2000). This can result from conflicting reports, but it can also be reported in the same publication. For example, Sattler et al. (1997) found that both, the CRKL-SH2 and the CRKL-SH3 domain, are able to bind CBL. In principle, the results regarding phosphotyrosine-dependency from the literature agree with the result from our screen. The interactions BMX-STAT3, SRC-PTK2 and NUMB-LNX2 have been described as dependent but were found as independent in this study. The first two are interactions involving kinases, and it is conceivable that the kinase activity of the bait or prey can make a phosphorylation-dependent interaction appear independent in the yeast two-hybrid system. The

latter is possibly classified incorrectly, or it is an example of a PTB domain binding in a phosphotyrosine-independent manner, as it has been suggested in (Rice et al. 2001). The interactions ABL2-CRK, CRK-CBLC, CRK-SOCS1 and PIK3R1-SOCS1 have been described as independent but were found as dependent in this study. Interestingly, three of these involve CRK. It is possible, that these are more examples of a simultaneous SH2 and SH3 binding. It is also possible to imagine a mechanism of intramolecular rearrangement allowing an SH3 domain-mediated interaction only upon phosphorylation. Of note, the interaction between CRKL and SOCS3 has been reported to be mediated by the CRKL SH2 domain (Sitko et al. 2004).
In conclusion, the comparison to the literature shows that, while the data set is of high quality and mostly new, the number of missed interactions is probably very high, as is common in protein-protein interaction detection experiments (Braun et al. 2009).

3.11 Subcellular protein interaction localization in intact cells reveals high flexibility of the GRB2 adaptor protein

To further raise the level of confidence in the phosphotyrosine-dependent protein-protein interaction set, especially in biological context, interactions between differentially located proteins were selected and visualized in intact mammalian cells using a split-YFP assay. To ward against faulty literature data, the localization information was controlled in immunofluorescence experiments using Protein A-tagged proteins and, where available, using antibodies against the endogenous proteins. Figure 23 shows very different localization of GRB2 protein interactions. GRB2 alone can be seen mainly in the cytoplasm (Figure 23 A and D). With RB1, which is nuclear (Figure 23 B and E), GRB2 interacts in the nucleus (Figure 23 H). With C10orf81, which can be found in the cytoplasm and in the nucleus (Figure 23 F), the GRB2 interaction can be seen in the cytoplasm (Figure 23I). The interaction with the membrane protein TSPAN2 (Figure 23 C and G) is found in membranous outgrowths induced by TSPAN2

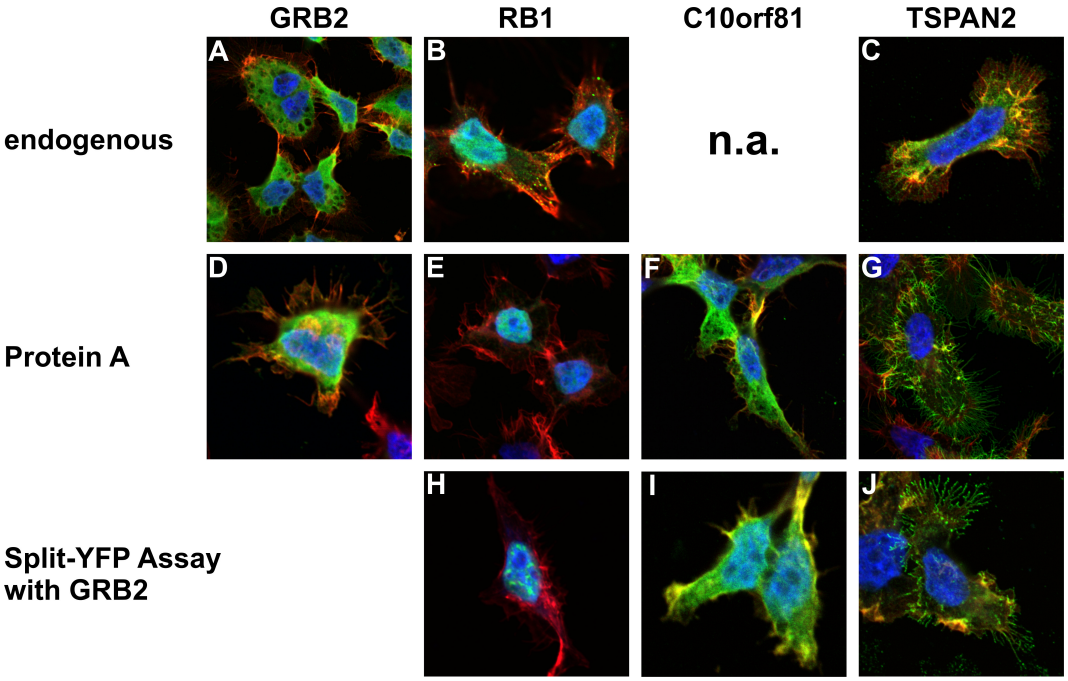


Figure 23: The adapter protein GRB2 shows high locational flexibility in its interactions. (A-C) Endogenous localization of GRB2, RB1 and TSPAN2, as determined by immunoflurescence using specific antibodies (green), (D-G) subcellular localization of Protein A-tagged GRB2, RB1, C10orf81 and TSPAN2, as determined by immunoflurescence using generic antibodies (green), (H-J) protein complementation assay results for the interactions GRB2-RB1 (H), GRB2-C10orf81 (I) and GRB2-TSPAN2 (J) (green), (all) phalloidin (red), nucleus/DAPI (blue).

(Figure 23 J). In agreement with its role as signaling hub, GRB2 is a prime example of a protein that can interact in different localizations, maybe triggered by the activity of different kinases. It should be noted that the protein complementation assay visualizes only the interactions. In other words, the overall distribution of the two proteins may remain unchanged. This would mean that the same protein can be present in several distinct complexes, possibly fulfilling different biological functions, simultaneously.

Chapter 4

Discussion

4.1 Yeast two-hybrid screening for phosphotyrosine-dependent protein interactions is the best way of improving our understanding of multicellular regulation of growth and development and related disease

4.1.1 Fast and efficient signaling is important for multicellularity

From an evolutionary perspective, signaling is an extremely important prerequisite for the development of complex systems. Biological entities form complex systems by building hypercycles (Eigen & Schuster 1977) that act as single evolutionary units. When multicellular lifeforms emerge, the single cells give up their independence and selection acts one organisms instead of single cells (Hoenigsberg et al. 2008; Smith & Szathmary 1997). Where cells start showing selfish behavior, the organism develops cancer and perishes. Once the transition from unicellular to multicellular organisms starts, the multicellular entities that display the most coordinated responses to the challenges they encounter have a selective advantage over those that do not. For an efficient coordination of cellular behavior on an organism/tissue scale, signaling mechanisms need to function in a highly effective, fast and independent manner.

4.1.2 Dedicated proteins with phosphotyrosine reader-, writer-, or eraser-function facilitate rapid and specific signaling

The use of dedicated signaling molecules achieves a high degree of independence from the regulated systems and has a high potential for effectiveness. Augmenting the established mechanisms of protein regulation by synthesis and degradation with mechanisms that employ ”stand-by” systems that need only be switched on or off enables more rapid responses. Second-messenger and post-translational protein modification systems are possible solutions meeting these criteria. Both of these solutions require three kinds of proteins to constitute marked improvements upon other signaling mechanisms: one type of protein to effect the activation of the system, one type of protein to reverse the effect of the first and one type of protein that can sense the state of the system and act accordingly. Post-translational modifications are distinguished from second-messenger systems by the fact that post-translational modifications are clearly attributed to a specific signaling context by virtue of the protein they are attached to. In the case of phosphotyrosine signaling, the close connection between kinase function and phosphotyrosine binding is reflected in the tight evolutionary link in the diversification of SH2 domains and protein tyrosine kinases (Liu et al. 2011a). It is assumed that the SH2 domain and the protein tyrosine phosphatases emerged first, and, after the appearance of protein tyrosine kinases, formed a complete ”toolkit”, consisting of ”writer”, ”reader” and ”eraser” of protein tyrosine phosphorylation (Lim & Pawson 2010; Pincus et al. 2008).

4.1.3 Evolution of metazoan complexity parallels phosphotyrosine signaling

In practice, the writer and eraser functions are often found in association with the reader function, either in the same amino acid chain or by virtue of adapter proteins. For example, in the case of phosphotyrosine signaling, most non-receptor tyrosine kinases have at least one SH2 or PTB domain or are closely associated with one or more adapter proteins (Hubbard & Till 2000;

Blume-Jensen & Hunter 2001; Manning et al. 2002b).

Conversely, phosphotyrosine signaling is one of the hallmarks of metazoan evolution (Müller 2001; Hunter 2009; Darnell 1997). This can also be seen in a correlation of the number of different tyrosine kinases and the "complexity" of an organism (Hunter & Cooper 1985). A similar connection has been shown for the number of SH2 domain protein families and the number of different cell types of an organism (Vogel & Chothia 2006).

The importance of phosphotyrosine signaling for multicellular organisms is further exemplified by the rapid evolutionary burst following the emergence of multicellularity. In yeast, there is only a single known SH2 domain-containing protein (Maclennan & Shaw 1993), whereas in higher eukaryotes the SH2 domain family is one of the largest protein domain families (Venter et al. 2001). Interestingly, the genome with the highest number of SH2 domains belongs to a unicellular organism, *Monosiga brevicollis* (Manning et al. 2008). This organism is thought to be closely related to the last unicellular ancestor of multicellular organisms (King et al. 2008). Nevertheless, *M. brevicollis* has only 20 SH2 protein families (Liu et al. 2011a).

In human and other mammals there are 38 SH2 domain protein families, thirty-four of which were already present before the split of echinoderms and chordates (Liu et al. 2011a). Pincus et al. (2008) analyzed 41 eukaryotic genomes and found significant co-evolution of SH2 domain proteins, protein tyrosine phosphatases and protein tyrosine kinases.

As a result of the emergence of wide-spread protein tyrosine phosphorylation, evolutionary pressure against deleterious phosphporylation events led to a reduction in the relative number of tyrosine residues with increasing complexity in metazoans (Tan et al. 2009b). This observation, in turn, lends credibility to the conjecture that phosphorylation events retained through evolution are mostly functional. The results of SH2 domain profiling experiments performed by Machida et al. (2003) who find that in their experiments "nearly all of the bands detected in the anti-pTyr blot are represented in the superimposed SH2 profile" further support this line of reasoning.

4.1.4 Phosphotyropsine signaling is intimately connected to cancer and other diseases

Because of the intimate connection between phosphotyrosine signaling and regulation of growth and development, there is a number of highly relevant diseases for which alteration of phosphotyrosine signaling is a central aspect. First and foremost, virtually all kinds of cancer are associated with mutations in phosphotyrosine signaling genes, especially kinases (Blume-Jensen & Hunter 2001; Gschwind et al. 2004; Futreal et al. 2004). The most relevant cancer pathways contain many genes that are also found in phosphotyrosine signaling pathways (Vogelstein & Kinzler 2004). Nevertheless, phosphotyrosine signaling has also been shown to play a pivotal role in a number of other diseases, like agammaglobulinemia and Noonan's syndrome (Yaffe 2002), dyslipidemia, hypertension, cardiovascular disease, stroke, blindness, kidney disease, female infertility, and neurodegeneration (White 2003), as well as hypercholesteremia, familial stroke, coronary artery disease, Alzheimer's disease, and diabetes (Uhlik et al. 2005).

4.1.5 Several, orthogonal methods have been used to investigate phosphotyrosine signaling

4.1.5.1 Peptide binding arrays efficiently assay preference for linear peptide motifs

SH2 domains often display a preference for certain amino acid residues in a fixed position toward the recognized phosphorylated tyrosine residue. These preferences are most noticeable for residues that are close to the phosphorylated tyrosine residue in the primary structure. Therefore, many SH2 domain binding preferences have been characterized in terms of short linear amino acid motifs using in-vitro peptide binding arrays using peptide arrays. There are two different approaches for peptide array binding experiments. The first one uses randomized oligopeptides (commonly 11-15-mers) that have a single fixed position in each spot. By probing each combination of amino acid residue and position and comparing relative binding signals, binding preference sequence motifs were determined for many SH2 domains with relative ease (e.g. Beebe

et al. 2000; Huang et al. 2008). However, these motifs assume an absence of cooperative effects, either positive or negative. They also fail when SH2 domains show multiple specificity, like the SH2 domain of PTPN11 (Wavreille et al. 2007). Empirically, the resulting motifs fit some protein-protein interactions very well, but fail at explaining others. Further, there are tyrosine motifs that match the consensus, but are not observed in the context of a full-length protein. Therefore, a second, more sophisticated, approach has been developed. Instead of randomized peptides, all (or the most relevant) naturally occurring tyrosine motifs can be used (e.g. Liu et al. 2012). This approach has the advantage that an observed binding event can be more easily used to derive a biological hypothesis, but it still measures peptide binding only. Another approach, that can be considered an intermediate, is a two-step peptide array system that determines permissive residues first, and then, in the second step, non-permissive residues, i.e. residues that abrogate binding (Liu et al. 2010).

4.1.5.2 Affinity purification-coupled mass spectrometry-based experiments inherently detect phosphorylation-dependent and -independent protein interactions

Another type of experiment that has proven fruitful not only in the field of phosphotyrosine binding analysis is affinity purification coupled to mass spectrometry. In experiments of this kind, affinity-tagged proteins are expressed in mammalian cells and enriched by binding to a chemical ligand or an antibody. Proteins bound to the tagged proteins are then identified by mass spectrometry, potentially allowing the detection of all proteins complexed with the protein of interest in the biological sample. In contrast to peptide array binding experiments, this method has several advantages. First and foremost, the binding event is observed in the context of complete proteins. Factors that are completely ignored in peptide-based experiments, like accessibility and protein conformation, have a normal influence in affinity purification experiments and even complex binding surfaces are intact. Furthermore, binding occurs, at least initially, in intact mammalian cells and post-translational modifications can be assumed to resemble biologically relevant situations. On the other hand, purified complexes have to be washed thoroughly before protein identification. Therefore, mass spectrometry-based methods are thought to have a bias toward

strong and stable interactions. Additionally, affinity purification-mass spectrometry interactions are not binary. Membership of the same complex does not necessarily imply physical interaction. Failure to account for these inherent differences leads to faulty analyses (Venkatesan et al. 2009).

While direct observation of phosphorylated peptide species is possible using mass spectrometric methods, special care must be taken in order to detect these signals. Therefore, most studies rely on secondary rationales, like comparing stimulated and unstimulated cells and reasoning that stimulation implies phosphorylation. Another disadvantage of mass spectrometry-based methods for the detection of phosphorylation-dependent protein interactions is that the observed effect can always be indirect and that there is always a background of kinase activity, that cannot be suppressed, because it is necessary for the survival of the mammalian cells.

In principle, dynamics of phosphorylation and protein binding are amenable for analysis by mass spectrometry-based methods. In practice, only a "countable number of studies have been reported on linking dynamics of protein interaction networks of protein complexes to cellular processes or signaling pathways using AP-QMS strategies" (Kaake et al. 2010). Most of the studies examining dynamics of phosphorylation or phosphorylation-dependent protein binding, rely on some kind of stimulating molecule related to the biological question under investigation. Proteins from stimulated cells are compared to proteins from unstimulated cells and differences are interpreted as consequences of the treatment (Pflieger et al. 2008; Rinner et al. 2007; von Kriegsheim et al. 2009; Olsen et al. 2006). Combination of growth factor stimulation and mutated bait proteins provide additional insight. For example, (Bisson et al. 2011) compare GRB2 complex composition for six different growth factor treatments using SH2 or SH3 binding-incompetent as well as wildtype constructs in AP/MS experiments. This approach allowed the characterization of 89 GRB2 complex binding partners in terms of SH2 or SH3 domain dependency. For 65 proteins binding depended on either domain. Signal strength for these interactions were compared after treatment with EGF, FGF, HGF, IGF, PDGF or insulin. Last but not least, the duration of EGF treatment was even varied from 1 through 100 minutes.

4.1.5.3 The yeast two-hybrid system relies on external tyrosine kinases for phosphorylation-dependent protein interaction detection

The yeast two-hybrid system is a well-tried method for the detection of protein-protein interactions. The two proteins of interest are fused to a transcription factors DNA binding or transcriptional activation domain, respectively. Interaction of the two proteins results in the expression of reporter genes fused to the corresponding transcription factor binding site. Accordingly, the yeast two-hybrid system produces binary interaction and is able to detect transient, "between-complex" interactions (Phizicky et al. 2003). It is very popular for proteome-wide screening because yeast two-hybrid are usually of high quality and it is amenable to automation. For example, Yu et al. (2008) reexamined about 100 interactions from the high throughput yeast data sets provided by Uetz et al. (2000) and Ito et al. (2001), the high throughput mass spectrometry data sets provided by Gavin et al. (2002) and Krogan et al. (2006), as well as literature-curated low throughput data, by yeast two-hybrid and YFP complementation and found that high throughput yeast two-hybrid data are of very high quality, when analyzed in the proper, binary context. Venkatesan et al. (2009) determined a precision of 79% for the yeast two-hybrid system in a proteome-wide context.
Because there is virtually no tyrosine kinase activity in *S. cerevisiae* (Gnad et al. 2009), phosphotyrosine-dependent interaction can only be detected where external kinase genes are introduced.

4.1.5.4 High throughput mass spectrometry continues to reveal vast numbers of protein phosphorylation sites

A question that is closely related to phosphotyrosine-dependent protein-protein interactions is which tyrosine residues are phosphorylated under which conditions. In some regards, this question can be regarded as basal to phosphorylation-dependent protein interaction research. Historically, the description of protein tyrosine phosphorylation laid the foundation for the discovery of phosphotyrosine-dependent protein-protein interactions (Sadowski et al. 1986). In a functional framing, the protein tyrosine phosphorylation is a prerequisite

for phosphotyrosine-dependent binding.

Protein tyrosine phosphorylation began receiving scientific attention after the description of SRC protein tyrosine kinase activity (Hunter & Sefton 1980) and the realization that this activity plays an important role in cellular transformation (Sefton et al. 1980). Yet, the amount of protein tyrosine phosphorylation is much lower than the amount of protein serine or protein threonine phosphorylation. Hunter & Sefton (1980) determined the relative amounts of protein serine, threonine and tyrosine phosphorylation as roughly 90%, 10%, and 0.05%, respectively. Olsen et al. (2006) measured relative amounts of 86.4%, 11.8%, and 1.8%, respectively. The striking difference in the proportion of tyrosine phosphorylation originates in the different methods. While the former used isotopic labeling to determine the relative number of actual phospho groups bound, the latter used mass spectrometry to identify unique sites. In other words, proteins carrying tyrosine phosphorylations are apparently less abundant than those carrying serine or threonine phosphorylations, on average. A comparison of the numbers of annotated phosphorylation sites comes out even more on the side of tyrosine phosphorylation due to research bias and owing to the high quality of pan-phosphotyrosine-specific antibodies. The first tyrosine phosphorylation sites were found by focusing on the targets of the first described protein tyrosine kinases, like the kinases themselves (Smart et al. 1981; Downward et al. 1984) or other, non-kinase proteins, such as ANXA2 (Radke et al. 1980). More recently, large-scale mass spectrometry-based studies have probed the phosphoproteome in an unbiased fashion and revealed large numbers of mostly novel phosphorylation sites (Beausoleil et al. 2004; Rush et al. 2005; Olsen et al. 2006). In 2009, Tan et al. (2009a) found 11,731 phosphoserine/-threonine and 9,283 phosphotyrosine sites in the publicly available resources. They determined that more than 65% of the conserved phosphosites were found in the 5 years prior. More recently, (Woodsmith et al. 2013) reported the number of phosphoserine/-threonine and phosphotyrosine sites as 56,251 and 13,241, respectively. In other words, in contrast to phosphotyrosine-recognizing domains, the number of reported phosphotyrosine sites is growing rapidly and does not show any indications of reaching saturation.

Therefore, a genome-scale screen focusing on phosphotyrosine-recognizing domain-containing proteins promised much more comprehensive results than one

based on known tyrosine-phosphorylated proteins.

4.1.5.5 Full-length protein context has been largely neglected in phosphotyrosine-dependent protein interaction investigation

The ever-growing number of protein tyrosine phosphorylation sites calls attention to the limited knowledge about phosphotyrosine binding. In light of the notion that the number of non-functional protein tyrosine phosphorylation events is expected to be small, it suggests that there are factors different from primary amino acid sequence contributing to phosphotyrosine binding specificity. An exact quantification of the impact the peptide motif sequence immediately surrounding the phosphorylated tyrosine residue has on binding is difficult. For phospholipase C γ, Bae et al. (2009) used isothermal titration calorimetry to show that binding to FGFR1 is much stronger than to the respective peptide alone and Min et al. (2009) demonstrated, that point mutations outside the peptide binding pocket diminish binding of ITK. For kinases, Linding et al. (2007) estimated the context to contribute at least 60% to substrate specificity. It seems likely that context contributes equally to protein-protein interactions. Therefore, the current knowledge of phosphotyrosine-dependent protein-protein interaction will profit greatly from an interaction map derived from experiments using full-length bait, prey and kinase proteins.

4.1.6 Genome-scale yeast two-hybrid screening of SH2 and PTB domain genes fills a gap in phosphotyrosine-dependent protein-protein interaction literature

4.1.6.1 Signaling proteins employ domains as molecular building blocks

The increase in complexity seen in eukaryotes goes hand in hand with higher modularity. In protein structure, relatively self-contained modules called domains evolved. Protein domains can usually fold independently (Bagowski et al. 2010) and tend to lie with single exons, especially in complexer organisms (Liu & Grigoriev 2004). Consequently, in protein evolution, structure is retained more

than amino acid sequence (Chothia & Lesk 1986). Most proteins comprise more than one domain and fulfill functions that are directly related to their domain composition (Han et al. 2007). Alternatively, adapter proteins are employed (Pawson & Scott 1997). Novel combinations of existing domains allowed for more complex organisms (Koonin et al. 2000). Indeed, this principle was even exploited very successfully in protein engineering. Desai et al. (1993) demonstrated that PTPRC-deficient Jurkat cells can be rescued by an EGFR-PTPRC chimera. Even more, EGF inhibits PTPRC function by inducing dimerization. Howard et al. (2003) created chimeric adaptors by fusing the GRB2 SH2 domain or SHC1 PTB domain to the death effector domain of FADD. These chimeric adaptors were able to induce caspase activation and cell death in response to mitogenic or transforming RTK signals. It has been observed early, that "[t]he main biological role of protein modules seems to be for specific protein-protein interactions." (Campbell & Baron 1991). While others (Yu & Lemmon 2003) report the percentage of human genes with protein interaction domains to be only 2.5%, the percentage of signaling genes is very similar and the two sets have a high overlap. Therefore, it can be safely assumed that most signaling protein interactions are domain-mediated. In fact, even in prokaryotes domain information can be used to improve protein interaction prediction significantly. This was demonstrated by Wojcik & Schächter (2001), who compared protein-protein interaction prediction for *Heliobacter pylori* based on *E. coli* interactions with and without domain information and benchmarked the results using the experimentally generated interaction map provided by Rain et al. (2001). Therefore, we followed a domain-driven approach in our analysis of tyrosine phosphorylation-dependent protein-protein interactions.

4.1.6.2 SH2 domains are highly relevant for phosphotyrosine signaling

Using prior knowledge as an indication, a relatively small number of SH2 domain-containing proteins was expected to account for the majority of the phosphotyrosine-dependent interaction data, but these estimates can be highly influenced by research bias. For example, GRB2 was first described in the context of phosphotyrosine signaling (Lowenstein et al. 1992) and has been scrutinized for phosphotyrosine binding since then while TP53 has been recognized as

an important cancer gene in 1979, but has not been researched in terms of phosphotyrosine signaling until the 1990s (Harris 1996). Furthermore, almost all SH2 domains are expected to bind to phosphotyrosine residues in proteins (Liu et al. 2006). Admittedly, there are reports of SH2 domains fulfilling other functions. For example, the SH2 domains of PI3K are able to bind PIP3, which competes for bindings with INSR (Rameh et al. 1995), the TENC1 SH2 domain binds DLC1 in a phosphorylation-independent manner (Dai et al. 2011), and the SH2 of JAK1 was suggested to function mainly as a peptide linker, because introducing a mutation known to disrupt phosphotyrosine recognition did not affect its subcellular localization or signaling capacity (Radtke et al. 2005). Nevertheless, these reports seem singular and do not necessarily preclude tyrosine recognition. Therefore, all human SH2 domains should be assayed in a comprehensive screen for phosphotyrosine-dependent protein-protein interactions.

4.1.6.3 SH2 domains are well-defined and have virtually no functions other than phosphotyrosine recognition

There is a large number of SH2 proteins in the human genome. This number of bona fide SH2 domains and SH2 proteins in the human genome has grown from 95 SH2 domains in 87 proteins (Venter et al. 2001) to 121 SH2 domains in 111 proteins (Liu et al. 2011a). Pawson et al. (Pawson et al. 2005) talk about 115 human SH2 domains. Jones et al. (Jones et al. 2006) give a slightly lower number of 108 SH2 domains in 98 proteins. In the same year, Liu et al. (2006) describe 120 SH2 domains in 110 proteins. In the next five years, this set was expanded only by SH2D7 (Liu et al. 2011a). Because of the intimate link to the regulation of processes necessary for multicellular organism and phosphotyrosine signaling, SH2 proteins can be categorized in eleven functional groups (Liu et al. 2006). Known disease-causing mutations in SH2 genes result mostly in various types of cancer or immune disease (Liu et al. 2006). Similarly, phenotypes observed in SH2 gene knock-out mice are mostly related immune functions (Liu et al. 2006). Thus, the group of SH2 domain-containing proteins is highly relevant for phosphotyrosine signaling, has virtually no other known functions than phosphotyrosine binding, is well-defined and important for pressing medical issues.

4.1.6.4 PTB domains have to be considered in a comprehensive analysis of phosphotyrosine signaling

The PTB domain-containing genes, on the other hand, are much more loosely defined than the SH2 domain-containing genes. The PTB domain has a PH fold (Blomberg et al. 1999), and the boundary between PTB and PH domains varies between different authors. Some authors consider the PTB domain a subclass of the PH domain (Rebecchi & Scarlata 1998), others regard it as separate (Uhlik et al. 2005). Even functionally the PTB domains cannot be distinguished sharply. There are subtypes, that clearly recognize to phosphorylated tyrosine residues in proteins and others for which this has not been shown (Uhlik et al. 2005). Even within the same subtype, there can be PTB domains that bind phosphotyrosine residues, and others that do not. For example, the PID or Dab-/Shc-like subtype contains domains that bind phosphotyrosine residues, like in Shc proteins, and others that bind to non-phosphorylated tyrosine residues or even phenylalanine residues, like the ones found in Dab proteins (Uhlik et al. 2005). An additional problem is that other functions, like the ability to bind to phosphate groups on non-protein molecules, are often carried by the same domains. For example, the Shc PTB domain can bind phosphatidylinositol-4,5-bisphosphate and phosphatidylinositol-4-phosphate (Ravichandran et al. 1997). Thus, the PTB domain-containing proteins are also relevant for phosphotyrosine signaling, but have many members that have different functions and are less well-defined than SH2 proteins.

4.1.6.5 The relevance of non-SH2, non-PTB domains for phosphotyrosine signaling seems negligible

In addition to SH2 and PTB domains, some PH domains not usually referred to as PTB domains have been shown to bind proteins on phosphotyrosine residues. However, the number of PH domain-containing genes is very high, totaling more than 600 genes, and the expected fraction of genes showing a phosphotyrosine binding ability is expected to be very low. There are even completely unrelated domains, like the C2 domain of PRKCD, that have been shown to recognize phosphotyrosine sites on proteins (Benes et al. 2005). Still,

the reports of unrelated domains binding phosphotyrosine residues seem singular and anecdotal.

4.1.6.6 All genes containing SH2, IRS1-type or PTB-type PTB domains define a comprehensive basis for phosphotyrosine-dependent protein interaction screening

The more clearly defined PTB domains appear to be the most relevant ones for phosphotyrosine signaling. Therefore, the screen was restricted to the PTB domain subgroups of the IRS1 and PID subtype. These subtypes are known to be substantially involved in phosphotyrosine binding. Additionally, all SH2 domain-containing genes were chosen as bait, even the few that were suspected not to bind phosphotyrosine residues. In practical terms, the InterPro identifiers IPR000980 (SH2), IPR002404 (PTB_IRS1) and IPR006020 (PTB_PID) were used to query the UniProt database and the resulting entries were mapped to Entrez Genes. We then attempted to obtain protein interactions for at least one clone belonging to each Gene ID, thus generating a comprehensive set of phosphotyrosine-dependent protein-protein interactions (Table 4).

4.1.6.7 Human non-recptor tyrosine kinases allow detection of phosphotyrosine-dependent protein-protein interactions in the yeast two-hybrid system

We wanted to gain insight into the organization of cellular signaling by tyrosine phosphorylation. Towards this end, we used the yeast two-hybrid technique to generate a network of phosphotyrosine-dependent protein-protein interactions, that is as complete and unbiased as possible. A set of bait proteins containing phosphotyrosine-recognizing domains was screened against a proteome-scale matrix of prey proteins in the presence of non-receptor tyrosine kinases.
Introducing a non-receptor tyrosine kinase allows yeast two-hybrid screening for phosphotyrosine-dependent protein-protein interactions. The kinase gene is encoded on a third plasmid. This makes it compatible to our established matrix-based yeast two-hybrid system (Stelzl et al. 2005), allowing genome-scale screening in a very sensitive, stringent, well-proven setup (Venkatesan et al. 2009). The system was used to screen a comprehensive set of phosphotyrosine-recognizing domain-containing genes. Bait strains with phosphotyrosine-

recognizing domain-containing genes and kinase plasmids were prepared and tested for autoactivation. Non-autoactive bait strains were then screened against a genome-scale prey matrix. Finally, phosphotyrosine-dependent interactions were distinguished in an independent retest experiment. The use of a retest experiment is an established concept in our lab (Hegele et al. 2012; Worseck et al. 2012), as well as in others (Goehler et al. 2004; Stelzl et al. 2005; Rual et al. 2005; Braun et al. 2009; Vinayagam et al. 2011). It makes the results more stringent because it uses fresh yeast and short growth period with a limited number of replications. In this study it has an additional purpose. Interactions were recognized as phosphotyrosine-dependent by comparing the same combination of bait and prey with different kinases and empty kinase vector controls. At the same time the retest provides an even higher degree of stringency than in conventional yeast two-hybrid. In addition to revealing phosphotyrosine-dependency, the empty kinase vector controls bait and prey autoactivity, because the kinase did not contribute to autoactivity. More than that, it did so in a way that is more relevant than in conventional yeast two-hybrid, because bait and prey are expressed in the same cell. This additional stringency allowed us to abstain from using the additional reporter gene read-out. To test this, we assayed 32 phosphotyrosine-dependent interactions that had already scored positive in the growth assay for β-galactosidase activity. This provided virtually no additional information.

There is virtually no background tyrosine kinase activity in *S. cerevisiae* (Gnad et al. 2009; Manning et al. 2002a), a fact that has been taken advantage of before, e.g. to investigate the suppression of SRC kinase activity by CSK (Murphy et al. 1993). Overexpression of the SRC tyrosine kinase has been reported to exert toxicity in *S. cerevisiae* (Brugge et al. 1987; Boschelli et al. 1993; Florio et al. 1994). On the other hand, the successful introduction of a tyrosine kinase into a yeast two-hybrid system has been reported several times in the past. Dombrosky-Ferlan & Corey (1997) used a yeast two-hybrid system to show that the association between the Src family kinase LYN and the regulatory subunit of the phosphatidylinositol 3-kinase PIK3R1 is mediated by the E3 ubiquitin ligase CBL and that the interaction between PIK3R1 and CBL depends on the presence of a kinase-active LYN kinase. Rocchi et al. (1998) used a plasmid expressing both, a human insulin receptor β(INSR) fragment and a

LexA fusion protein, to show that PIK3R1 and PTPN11 interact with GAB1 in a phosphorylation-dependent manner. Delahaye et al. (2000) used the same system to screen a human placenta cDNA library and successfully identified FRS2 as a phosphotyrosine-dependent interaction partner of PTPN11. Yamada et al. (2001) report a GAL4-based yeast two-hybrid system employing a third plasmid expressing a rat Ntrk2 gene fused to a glutathione-S-transferase gene to screen a commercial human brain cDNA library. Additionally, they found phosphorylation-dependent interactions of PTPN11 with GRB2 and SIRPA. There is a number of other yeast two-hybrid studies that, with or without intent, used protein tyrosine kinases to detect phosphotyrosine-dependent protein interactions (Keegan & Cooper 1996; Osborne et al. 1995; Marti et al. 1998; Warner et al. 2000; Cao et al. 2002; Sayós et al. 2004; Ingley et al. 2006; Sylvester et al. 2010). Nevertheless, all of these are centered on one or a few genes. In this study, we present a systematic yeast two-hybrid screen for phosphotyrosine-dependent protein-protein interactions.

4.1.6.8 Yeast two-hybrid provides high-quality, binary protein-protein interactions

In an attempt to create a systematic and comprehensive set of phosphotyrosine protein-protein interactions, we used a yeast two-hybrid system for genome-scale screening. Yeast two-hybrid has several advantages over other methods for protein-protein interaction detection.
It provides binary, physical interaction data. This means that for every interaction, there are only two interacting proteins and that they come into close proximity to interact directly with each other. It uses full-length proteins and, last but not least, has an extremely low false positive rate. A low false positive rate is especially important for screening applications, because of the comparatively small fraction of positives. The false discovery rate of an interaction detection experiment is given by the formula

$$FDR = FP/(TP + FP),$$

with FP and TP defined as the number of false and true positives, respectively. The two can be calculated by multiplying the respective rate, the number of

tests and the prior propabilities. Thus,

$$FP = n * FPR * (1 - PR_{prior}),$$

with n, the number of tests, FPR, the false positive rate and PR_{prior}, the expected positve rate. Similarly,

$$TP = n * (1 - FNR) * PR_{prior},$$

with FNR, the false negative rate. In unbiased proteome-wide screens, PR_{prior} is extremely low. Therefore, the false discovery rate may become substantial even for relatively low false positive rates. Braun et al. (2009) benchmarked five binary interaction detection methods and found that the false positive rate of yeast two-hybrid is substantially lower than that of all other methods. Venkatesan et al. (2009) estimated that they found about eight false positives in every 1,000,000 interactions assayed.

4.1.6.9 Systematic yeast two-hybrid data complements phosphotyrosine-dependent protein interaction knowledge

Methods for the identification of protein-protein interactions are generally limited in sensitivity and are orthogonal (Braun et al. 2009; Venkatesan et al. 2009). This means that each methods identifies a subset of the interaction in the assayed interaction space only and that the subset of interactions that can be found with each method appears to be a different and independent draw from a pool of true interactions. Consequently, the best approximation of a complete interactome map is obtained by combining the interactions obtained by different detection methods. To our knowledge, there is a multitude of studies using tyrosine kinases in low throughput yeast two-hybrid systems (Keegan & Cooper 1996; Dombrosky-Ferlan & Corey 1997; Osborne et al. 1995; Rocchi et al. 1998; Marti et al. 1998; Delahaye et al. 2000; Warner et al. 2000; Yamada et al. 2001; Cao et al. 2002; Sayós et al. 2004; Ingley et al. 2006; Sylvester et al. 2010) or *E. coli* -based two-hybrid systems on the one hand (Shaywitz et al. 2000), and a number of high throughput studies investigating phosphotyrosine-dependent protein-protein interaction with mass spectrometry-based (Schulze et al. 2005; Blagoev et al. 2003; Oyama et al. 2009; Schulze

& Mann 2004) or other methods (Yaoi et al. 2006; Miller et al. 2008; Jones et al. 2006; Li et al. 2008) on the other hand, but no high throughput yeast two-hybrid study. This is astonishing given the inherent advantages of this method, like the full-length protein context, the binary interaction results, the clear kinase-interaction relationships and the virtually non-existing tyrosine phosphorylation background. The selection of yeast two-hybrid as the main method for this study is therefore not only prudent, but it also fills a void in the current phosphotyrosine-dependent protein-protein interaction knowledge.

4.2 State-of-the-art yeast two-hybrid screens are the technical basis for a high-quality phosphotyrosine-dependent interaction data set of considerable size

4.2.1 Comprehensive sets of bait and kinase genes and a genome-scale prey matrix facilitate unbiased and complete results

In an attempt to provide a comprehensive and unbiased network of phosphotyrosine-dependent protein-protein interactions a set of genes containing phosphotyrosine-recognizing domains was co-transformed with a representative set on non-receptor tyrosine kinases and screened against a genome-scale prey matrix.

For twenty-two, so about fifteen percent, of the target genes, no clone was obtained.

The list of genes and the number of clones obtained and screened is shown in Table 4. There are 149 target genes, 110 of which have at least one SH2 domain and 46 of which carry at least one PTB domain. Seven genes have both.

The bait-kinase combinations were arrayed in pools that fit into microplates. This allowed relatively easy preparation and propagation of pools for autoactivation testing, screening and retesting. For the autoactivation test, the bait-kinase pools were mated with an empty prey strain and controlled for growth.

For screening, the pools were grown directly in deepwell plates, without any re-arraying, leaving autoactive positions non-inoculated. By virtue of this practice, the yeast growth periods were kept as short as possible, reducing the incidence of late autoactivation, thereby raising the quality of the experimental results.

4.2.1.1 Autoactivation test and retesting guarantee high quality interaction data

In our yeast two-hybrid screens, all bait strains were tested for the ability to induce prey-independent growth on -HAULT medium. Such bait strains were rigorously excluded from further experiments. Furthermore, every positive interaction was reproduced in an independent experiment with fresh yeast. This practice guarantees high quality of the resulting protein-protein interaction data (Vinayagam et al. 2011)

4.2.1.2 Repeating the screening experiments with independently produced yeast strains reduces the number of undetected protein interactions

4.2.1.2.1 Repeat screening raises the chance of detecting weak interactions

Given the yeast two-hybrid systems inherent tendency for showing false-negative interactions in an apparently transformation-dependent manner, repeating the screens with several independently transformed bait strains raises the coverage of the assayed interaction space drastically (Venkatesan et al. 2009; Worseck et al. 2012). At the same time, it allows filtering of possible interacting prey constructs before the retest by the number of times a prey construct is found in repeated experiments. This practice reveals substantially more interactions with the same effort (Venkatesan et al. 2009; Worseck et al. 2012).

4.2.1.2.2 Repeat screening reduces the number of clones excluded due to occasional autoactivity

The number of genes with autoactive clones (i.e. clones that needed to be removed because they produce a signal in the absence of a prey protein) is only

five, which is about 3% of all target genes. Considering that about 20% of all clones are expected to be autoactive in an unbiased setup, this number is very low (Stelzl et al. 2005; Goehler et al. 2004; Nakayama et al. 2002; Walhout & Vidal 2001). Apparently some clones are always autoactive, some clones are never autoactive and some clones show autoactive behavior depending on the yeast transformation.
For clones whose autoactivity depends on the transformation the four replicate screens and the many different transformations with the different kinases provide a sufficiently large number of transformations to guarantee that they can be screened. This reduced the rate of autoactives by a factor of about two, leaving twenty-five autoactive bait genes.

4.2.1.3 Our two-step screening protocol boosts coverage substantially

4.2.1.3.1 Testing extra clones for autoactive genes raises the number of genes included in the screening experiments

For clones that appear to be autoactive in every experiment, additional clones for the same gene were selected and tested in the second part of our two-step screening approach. This led to an additional reduction by a factor of five, leaving only five genes without non-autoactive clones.

4.2.1.3.2 Testing extra clones for initially unsuccessful genes raised the fraction of successful genes

We also included genes in the second screen that were not autoactive in the first screen, but failed to produce interactions. Therefore, the remaining genes have a high coverage in terms of clones that produced interactions when compared to other proteome-scale interaction studies. There are only 42% of the screened bait genes for which no clone produced any interactions. In the systematic matrix-based yeast two-hybrid screens reported by Stelzl et al. (2005) and Rual et al. (2005), this fraction is 76% and 87%, respectively. The reason for this is partly that the number of clones without interactions is low to begin with and partly the two-step screening process (i.e. the inclusion of additional clones for unsuccessful genes in a second screening round). Initially, there were

68 genes of the 121 screened ones without interactions, which is about 56%. By screening additional clones for 42 of these genes, this number was reduced by 17 genes (14%).

4.2.1.3.3 Benefits of the two-step screening protocol far outweigh potential biases

The two-step screening approach raises the fraction of genes with productive genes and reduces the percentage of genes unamenable for yeast two-hybrid screening because of autoactivity. Nevertheless, there are some drawbacks to this approach. The different treatment of genes can be a source of bias potentially making further analysis difficult:

1. The number of prey genes in the matrix was higher in the second round, because those prey genes were not available for the first round.

2. The number of kinases was lower, because it had become apparent that most of the kinases resulted none or very few interactions and the workload reduction was substantial.

3. Because there was no third round, for all genes that did not have bait clones producing interactions in the first round, all available clones were used in the second round, instead of just one, like in the first round.

The first point can lead to a greater number of interactions for genes screened in the second round and means that certain prey genes can only interact with bait genes screened in the second round. This could be cured by disregarding all interactions found with any of the additional prey genes. Since protein-protein interaction detection assays are plagued by large false negative rates (Braun et al. 2009), and, thus, can by no means claim absolute completeness anyway, the loss of interaction information is much graver than the possible bias for most analyses. The bias introduced by the second point is probably extremely weak, from what can be estimated from the results of the first assay. It can theoretically be cured by removing all interactions from the first round that did not retest positive with any kinase used for the second round. Since bait genes and kinases were pooled in the primary screen, it is conceivable that some interactions that retested positive with kinases used in the both rounds was

initially detected only because of the presence of one of the other kinases, but this is extremely unlikely. The third point cannot be cured, except by regarding the two rounds as independent screens, and can be thought of, at least partly, as an extension of the biases present in the availability of ORF clones. The bias that arises from screening a single additional clone for genes that had an autoactive or non-productive clone in the first round is hard to gauge. There is no obvious connection between non-productive and productive clones for a gene. In other words, the prior expectation in terms of, for example, degree does not change based on the existence of non-productive clones for the same genes. Similarly, there is no obvious rule for non-autoactive clones for genes that have autoactive clones, although, in this case, the probability that the next clone is also autoactive is higher, as autoactivity often seems to be associated with certain domains or amino acid sequences. If there is no connection between a genes degree and its clones propensity to be autoactive or non-productive, the practice of screening additional clones until a non-autoactive and productive one has been found does not create any bias. Even if there is a connection and a small biases arises from this method, the completeness gained outweighs the bias by far. A somewhat greater bias arises from situations where there is no successful clone in the first round but several successful clones in the second round. This aspect can be cured by randomly choosing one productive clone for each such gene and discarding all interactions not found with this clone. Again, the bias is weak, and the additional interaction knowledge outweighs it by far. To sum up, there may be some biases introduced by the two-step screening process, but they are probably very small and the benefits are substantial. For applications with very strict and specific requirements in terms of freedom from any bias, many can be cured at the cost of losing interactions. For most applications, this should not be necessary, because methods used to analyze protein-protein interaction data should be robust against missing interactions. Many analyses can be controlled internally. For example, the independent interaction set can be used for comparisons, because any assay-dependent bias should be similar to the phosphorylation-dependent interaction set.
In total, the two-step screening strategy seems like a good compromise between comprehensiveness and freedom from bias, making about half of the genes in the target set successful bait genes.

4.2.2 The discovered phosphotyrosine-dependent protein-protein interactions achieve high coverage in terms of width and depth

4.2.2.1 The number of discovered phosphotyrosine-dependent protein-protein interactions is considerable

Screen size can be measured in terms of the size of the interaction space and in terms of depth. The interaction space is given by the number interactions that could have been detected. In a binary screen, using the yeast two-hybrid system, this is the product of the number of bait genes and the number of prey genes. The screens depth is the number of interactions per bait gene or per interaction in the search space. Both can be described best in contrast to other matrix-based proteome-scale studies. This study has a much smaller number of bait genes than prey genes, which makes it different from the Stelzl et al. (2005) and the Rual et al. (2005) studies, which had roughly similar numbers of bait and prey genes. Consequently, the number of bait genes is much smaller than in the other studies while the number of prey genes is somewhat higher. The number of interactions found in this study is about half the number found in each of the other two. Because the number of bait genes is so much smaller than in the other studies, the number of interactions per bait gene is about three-fold higher than in each of the other studies. This study has 628 interactions found with 70 bait genes, which comes out to 9 interactions per bait gene. The Stelzl and Rual data sets have 3269 interactions for 1064 bait genes and 2754 interactions for 926 bait genes, coming out to 3.1 and 3 interactions per bait gene, respectively. Due to the different number of prey genes, the numbers are more similar for the number of interactions per possible interaction. Dividing the 9 interactions per bait by 17007 (the number of prey genes), puts the positive rate at 0.053%, which is very similar to the rate of 0.055% obtained by dividing the 3.1 interactions per bait gene for the Stelzl study by 5632, the number of prey genes in the Stelzl study. The ratio for the Rual study is 3 divided by 7195 prey genes, which comes out to 0.041%, so slightly lower.

4.2.2.2 Coverage in terms of false-negative rate is typical for protein-protein interaction screens

The coverage of the screen has been assesses by comparing the interactions found among bait genes to the interactions reported in literature for the same interaction space. This analysis puts the estimated false-negative rate at almost 90%, as is typical for protein interaction screens (Braun et al. 2009).

4.2.2.3 Comparison to phosphotyrosine interaction literature indicates that the size of the phosphotyrosine-dependent interactome may have been underestimated

The combined knowledge covers about 30% of all theoretically detectable interactions in the bait-bait interaction space. The bait-bait interactions found in this study are reflected substantially better in the literature than the other interactions. This reflects a general problem with literature data sets. Literature data sets are sociologically biased (Venkatesan et al. 2009). Additionally, protein-protein interaction literature databases may contain incorrect entries and should be attributed no higher confidence than experimental data. Of the 247 interactions curated for this study, only about two thirds (167) proved correct. This value is similar to previous analyses, like (Venkatesan et al. 2009) and (Cusick et al. 2009), who report fractions of 62% and 65%, respectively. Likewise, in contrast to prevalent assumption (Stumpf et al. 2008; Reguly et al. 2006; von Mering et al. 2002; Batada et al. 2006), high throughput studies have been shown to produce more reliable results than low throughput studies (Collins et al. 2007; Venkatesan et al. 2009; Cusick et al. 2009). Nevertheless, if we ignore these problems and assume that the success rate in our study is constant, we can estimate the size of the phosphotyrosine-dependent protein-protein interactome. To extrapolate the false-negative rate estimate from the bait-bait interaction space to the genome-scale interaction space, we can first calculate the effective false-negative rate from number of detected interactions (81) and the estimated total number of interactions (581) in the bait-bait space. Because the interactions in the bait-bait space have been assayed in bait-prey and prey-bait direction, the resulting false-negative rate of 89.8% is the square of the false-negative rate for the bait-non-bait interactions. The estimated size of the phosphotyrosine-

dependent protein-protein interactome can be calculated by dividing the number of phosphotyrosine-dependent interactions by one minus the false-negative rate. Following this reasoning, the human phosphotyrosine-dependent protein-protein interactome contains about 5,500 interactions. On the one hand, this value seems relatively high compared to the complete interactome, which is estimated to contain about 74,000-200,000 interactions (Venkatesan et al. 2009). This means that the phosphotyrosine-dependent interactions make up about 2.5-7% of the human interactome. On the other hand, it lends credence to the prognosis that "[t]he current catalog of characterized tyrosine phosphorylated protein, however, appears to represent a significant underestimate of the prevalence of tyrosine phosphorylation." (Hunter 2009).

4.2.2.4 The phylogenetic tree of SH2 domains is well-covered by the phosphotyrosine-dependent proteins interactions discovered

Apparently the phoshotyrosine-dependent and independent interaction data sets are not strongly related in terms of numbers of interactions per gene. The high coverage of the phylogeny given the fact that only half of the bait genes produced interactions is owed mainly to the high density of the network. A minor contributing factor is that the PTB domain-containing genes are slightly over-represented in the set of genes for which no clones have been screened. When projecting the interactions onto the phylogenetic tree of SH2 domains (Figure 10), the interactions appear spread more of less evenly. This is true for phoshotyrosine-dependent and independent interactions.

The distribution of interactions among the phylogeny agrees very well with prior expectations even in terms of numbers of interactions. The genes PIK3R3, CRK, STAT3 and GRB2 are among the genes with the most interactions in this study and among the genes with the highest number of phosphotyrosine-dependent interactions in the literature.

There are two clades without interactions, the one of VAV1, VAV2 and VAV3 and the one of SHC1, SHC2, SHC3 and SHC4. The first one might be explained by the relatively large size of the proteins, all of which contain more than 800 amino acid residues. Proteins of this size are known to work badly in protein-protein detection assays. In fact, Nakayama et al. (2002) even developed a yeast two-hybrid approach designed specifically to deal with this problem.

4.2.2.5 Experimental validation in mammalian cell culture affirms high quality

The quality of the protein-protein interactions found in this study has been assessed experimentally by transferring about 10% of the interactions to human cell culture and testing them by co-immunoprecipitation. The obtained validation rates are indistinguishable from validation rates obtained for positive reference sets (Braun et al. 2009), supporting a high confidence in our results.

4.3 The phosphotyrosine-dependent interaction network grants insight on system and gene level

The graphical representation of the phosphotyrosine-dependent network paints it as a typical protein-protein interaction network, but one that is very dense and focused on the set of target genes. This impression is generally supported by the topological parameters.

4.3.1 Statistical network analysis shows that the tyrosine-dependent interaction network is scale-free, unusually dense and build around a central core

4.3.1.1 Selection of appropriate data sets for comparison is important for a meaningful statistical network analysis of real-world networks

The selection of an appropriate frame of reference is of paramount important for any statistical evaluation. For protein-protein networks, the most meaningful conclusions are derived from comparison to another network that is as similar to the network examined, except in the aspect under scrutiny. As a consequence, the best choice may vary and an optimal choice may be unavailable.

4.3.1.1.1 The interaction data sets provided by Stelzl et al. (2005) and Rual et al. (2005) are the largest sets of binary interactions among human proteins

The meaning of the statistical network parameters is understood best in comparison to suitable reference data sets. The data sets that are used most often in this context are two matrix-based genome-scale yeast two-hybrid data sets presented by Stelzl et al. (2005) and Rual et al. (2005). These data sets are the largest binary human protein-protein interaction data sets available.

4.3.1.1.2 The interaction data sets provided by Lim et al. (2006) and Wong et al. (2007) are smaller but have an asymmetrical setup

Both of the usual reference sets used for matrix-type yeast two-hybrid screens, the Stelzl and Rual datasets, are symmetrical, i.e. they generally used each tested ORF as bait and as prey. In contrast, this study focused on a set of target genes, carrying a phosphotyrosine-recognizing domain, and screened this set of genes against a genome-scale prey matrix. It is conceivable that this fact alone has an impact on the network structure. For example, the bait genes (the target genes) might have a higher average number of interactions, simply because they were tested with many more potential partners than the other proteins. This is exactly what we observe. The average number of neighbors is about 9.9 for the bait genes and only about 1.6 for the other genes. Since pairs of non-target genes are not tested for interaction, the clustering coefficient is expected to be lower than for non-targeted approaches. Similarly, a targeted approach is expected to produce data sets with a higher degree of centralization.
Therefore we also compare our results to two data sets resulting from targeted matrix-based yeast two-hybrid approaches. The first one, presented by Lim et al. (2006), focuses on genes that either have Ataxia-causing mutations or are known to interact with such genes. It was generated using a setup similar to the study of Rual et al. (2005) with 54 bait genes. The data set was validated by GST-pulldown. Interestingly, there are only twenty-one bait-bait interactions in the data set, although thirty-one of the target genes were chosen because they are known interact with the other twenty-three. The other one (Wong et al.

2007) focuses on genes related to mitotic spindle formation in *S. cerevisiae*. The screening setup is very similar to this study. It has 102 bait-bait interactions. These two targeted studies may be more suitable reference sets than the two non-targeted ones.

4.3.1.1.3 Domain dependence places phosphotyrosine recognition between symmetrical and asymmetrical setup logic

On the other hand, if we consider the network of phosphotyrosine-dependent interactions alone and assume that the number of phosphotyrosine-dependent interactions that are not phosphotyrosine-recognizing domain-mediated is negligible, we can argue that the only bias introduced by the targeted approach is that bait-bait interactions are assayed twice as often as other interactions. In this case, the Stelzl et al. (2005) and Rual et al. (2005) data sets are appropriate references and the differences we observe are biologically relevant.

4.3.1.1.4 Intermediate setup symmetry suggests using all four reference data sets for statistical network analysis

To sum up, we use an asymmetrical screening setup, because we focus on phosphotyrosine-recognizing domain-containing genes as bait. This means that the largest reference data sets might be unsuitable. On the other hand, the best asymmetrical reference data sets are smaller and both of them have a certain biological focus that might influence our analyses. Additionally, we also consider the bait-bait interaction space alone, which is symmetrical. Therefore, there is no single best reference set and we compare our data to four sets total, two symmetrical and two asymmetrical ones.

4.3.1.2 Statistical network parameters are consistent with an unusually dense protein-protein interaction networks

4.3.1.2.1 High clustering coefficient and centralization suggest high density or modularity

The network of phosphotyrosine-dependent protein-protein interactions discovered in this study has a clustering coefficient of 0.027 and a centralization of

0.22. These numbers change only marginally when the independent interactions are also considered.
The clustering coefficient is higher than that of the Wong and even the Stelzl data set, which is unexpected because of the targeted setup. The extremely high clustering of the Lim data set is probably a direct consequence of the selection of target genes that are related by protein interaction. Like the phosphotyrosine-dependent network, the Wong data set is very coherent and contains many more bait-bait interactions than expected by chance. This means that it provides a good frame of reference for biologically coherent protein interaction data sets generated by targeted yeast two-hybrid approaches, which show higher than random clustering. The centralization of the phosphotyrosine-dependent network is much higher than for the Rual and Stelzl and also the Wong data set. It is comparable to that of the Lim data set. A higher degree of centralization is expected for targeted approaches, but is expected to be accompanied by low clustering. The fact that both, the clustering coefficient and the centralization, are relatively high, suggests that the network displays tendencies of a dense modular or hierarchical structure.

4.3.1.2.2 The unexpectedly high number of interactions among SH2 genes indicates biological coherence

The bait genes clearly favor other bait genes as interaction partners which means that the domain-containing genes are themselves highly tyrosine-phosphorylated. They are not necessarily phosphorylated to a higher degree than other proteins, but, if they are not, their tyrosine-phosphorylations are at least more meaningful for protein-protein interactions. This indicates that phosphotyrosine-dependent protein-protein interactions are coherent in terms of biological function and probably also evolutionary origin and development.

4.3.1.2.3 The arrangement of the phophotyrosine-dependent interactions fits a central computational function

Despite its high density and coherence the phosphotyrosine-dependent interaction network is open to influences from different fields and pathways. Since the main topic associated with tyrosine phosphorylation is regulation of

growth and development it needs to have connections to proteins associated with related functions. The network of phosphotyrosine-dependent interactions has to be understood as a computational control module integrating all the relevant signals into a compound cellular decision. A solution using several distinct modules or other functional modules using tyrosine phosphorylation for the regulation of different biological processes would also have been imaginable.

4.3.1.2.4 The phosphotyrosine-dependent protein interaction network is scale-free and has an unusual exponent

The phosphotyrosine-dependent interaction network is scale-free (Albert & Barabási 2002), as are virtually all protein-protein interaction networks (Barabási & Oltvai 2004). Interestingly, the ratio of high-degree to low-degree nodes is uncharacteristically high, as is the resulting exponent of -1.41 in the log-log-fitted degree distribution. This exponent usually lies between -2 and -3 for protein-protein interaction networks (Yook et al. 2004). The biological meaning of this finding is not immediately clear. It is possible that is purely coincidental or a consequence of the depth of the screen or the fact that the bait genes were not chosen randomly, but, at least implicitly, by biological function.

4.3.1.2.5 The unusual scale-free exponent may be related to additional controllability due to phosphorylation

It is also possible that the reason lies in the possibility of switching phosphorylation-dependent interactions on or off. This means that, at each time, or for each state, etc. only a subset of the interactions in the network are active, reducing the networks factual density. In this context, a network-theoretical study presented by Liu et al. (2011b) might be instructive. The authors argue that network controllability is highly related to the exponent, and that mediocre, practically useful controllability can only be achieved between exponents of -2 and -3, which is precisely what is usually observed for protein-protein interaction networks. In other words, the phosphorylation aspect might add an additional layer of regulation that needs to be resolved before the network can be used for classical, i.e. static, network analysis.

4.3.1.3 The phosphotyrosine-dependent interaction network has a central core

4.3.1.3.1 Most core genes discovered are known to be central to phosphotyrosine signaling

Although the fact that the network is organized around a core of proteins that keep the network connected is evident, the eight nodes forming the core in the network may not be the most important ones in the true cellular context, rather arising only from idiosyncrasies of the yeast two-hybrid screen. On the other hand, many of the core nodes are excellent candidates for this role. GRB2, CRK, CBL and SH2D2A are central adapters binding directly or indirectly to a magnitude of growth factor receptors, linking them to intracellular signaling networks (Dikic et al. 2003; Acuto et al. 2008; Olayioye et al. 2000; Oda et al. 2005; Normanno et al. 2006). PIK3R3 is an important player in the extracellular signal-PTEN-AKT1-MTOR axis (LoPiccolo et al. 2008) and SOCS4 and STAT3 are important in JAK/STAT signaling (Naka et al. 1999; Croker et al. 2008; Kile et al. 2001). Only APPL1 is not immediately recognized as an important phosphotyrosine-related signaling molecule. On closer inspection, it may still fit this role very well. It binds to ADIPOQ receptors (Mao et al. 2006), as well as molecules in both the PTEN-PI3K-AKT1-MTOR (Mitsuuchi et al. 1999) and the Rab/Gab (Mao et al. 2006; Lee et al. 2011) axes of growth factor action, regulates phospholipase C (Deepa et al. 2011) and EGFR (Lee et al. 2011) and has even been found in the nucleus (Miaczynska et al. 2004).

4.3.1.3.2 Network analysis identifies PTK2, OLIG1 and LNX1 as important phosphotyrosine signaling genes without SH2 or PTK domain

Another interesting question is whether there are any genes belonging to the set of highly related genes encompassing the bait genes. In other words, if the target genes had been chosen for network coherence, are there any genes in the network that should have been target genes and are they connected by biological function? These genes should be the most prominent genes in the network that are not bait genes. The non-bait gene of the highest degree is PTK2 with six interactions, followed by OLIG1 with five and LNX1 with four

interactions. PTK2 is highly related to tyrosine phosphorylation and cellular proliferation and migration (Hanks et al. 2003). It is a tyrosine kinase itself and extremely important in integrin-directed growth. OLIG1 is a transcription factor whose expression is correlated with cell survival in lung cancer (Brena et al. 2007). LNX1 is an E3 ubiquitin ligase specialized in mediating the proteasomal degradation primarily of PTB domain-containing proteins like NUMB (Nie et al. 2002). Its down-regulation leads to cell cycle arrest (Zheng et al. 2011).

4.3.2 The phosphotyrosine-dependent protein-protein interactions confirm and improve upon established phosphotyrosine signaling knowledge

4.3.2.1 Tyrosine phosphorylation-associated functions are strongly connected

The network of the phosphotyrosine-dependent interaction is very dense (i.e. the proteins are very highly connected) and consists of a single module. This suggests a strong connection between tyrosine phosphorylation and protein phosphotyrosine binding as a biochemical mechanism and the associated biological functions, like regulation of growth and development or immune response. Furthermore, all of these functions are linked by phosphotyrosine-dependent protein-protein interactions. This contrasts a notion of several independent, isolated modules employing tyrosine phosphorylation as a mechanism, one of several phosphotyrosine modules linked by other mechanisms or even one of many structurally related proteins functioning in different biological contexts, like zinc finger proteins (Laity et al. 2001). Therefore, biological functions regulated by tyrosine phosphorylation are probably overlapping. Since growth and development are often hard to distinguish clearly, this is comprehensible. The same is true for immune responses, many of which comprise a developmental process, termed maturation, followed by a strong replication of the adult immune cells adequate for the respective stimulus (Paul 2008). Thus, the findings presented here link to the question: How do cells integrate signals?

4.3.2.2 The phosphotyrosine-dependent protein interactions fulfill cellular signal integration functions

The question "How do cells integrate signals?" was first raised more than twenty years ago, when it was realized that different growth factor signaling pathways utilize the same signals and signaling molecules to affect different cellular responses (reviewed in (Chao 1992)), and still has not been answered to full satisfaction. (Schlessinger & Ullrich 1992) propose that each cell expresses a tissue-specific set of SH2 domain-containing proteins that result in different tissue-specific responses to the same growth factor. (Tan & Kim 1999) discuss four different hypotheses for tissue-specificity in RAS-MAP signaling. (Bhalla & Iyengar 1999) argue that signaling pathways overlap to form signaling networks that have properties that the single components do not have. In most of these models, there is a need for one or more mechanisms for separating the network components, either temporally or spatially.

The key to understanding signaling integration lies in protein phosphorylation as an additional layer of regulation. Signaling proteins can be present in the inactive, unphosphosphorylated form and become phosphorylated at the appropriate time and location only. As a consequence, the interaction maps are more dense than expected for biological networks based on controllability considerations (Liu et al. 2011b) and genes regulated by phosphorylation, like cancer genes, have a higher average degree and are more central than average (Jonsson & Bates 2006).

The additional layer of network control is governed by kinases and phosphatases, that effect and reverse phosphorylation, respectively. Both, kinases and phosphatases, are subject to regulation themselves. As demonstrated by the kinase plate assay experiments conducted in this study, there is a specificity regarding kinases and interactions that suggests the presence of several potentially phosphorylated tyrosine residues on most tyrosine-phosphorylated proteins, allowing for signal adjustment. This agrees well with findings regarding well-studied examples, like growth factor receptors or IRS1 and IRS2, that have been shown to have many potential phosphotyrosine residues and often have many-many relationships between binding partners and potential phosphotyrosine residues (Schulze et al. 2005; Hanke & Mann 2009). Integration of signals in vivo could thus be mediated by several kinases acting on a single protein with

several potentially phosphorylated tyrosine residues. The combined efficacy of several kinases might be a result of the single kinases binding affinities and phosphorylation efficacies on a single residue basis. Globally, kinases could act partially redundant and partially complementary to achieve complex regulatory logic. Such a configuration can achieve virtually any kind of signal integration logic, especially in combination with a similarly complex system of partially overlapping binding specificity in the reader components, i.e. the SH2 proteins. For instance, CBL is phosphorylated by the kinase domains of FYN, SYK and Abl in in-vitro assays, and each of these displays preferences for different tyrosine residues (Grossmann et al. 2004). Further, CBL can bind and ubiquitinate EGFR either directly or by virtue of adapter molecules. Interestingly, the ubiqitination has different functional consequences depending on the mode of binding, which has been demonstrated by means of EGFR carrying a point mutation on tyrosine 1045, the CBL binding site, and double-SH3-mutant GRB2 (Grøvdal et al. 2004). Similar arguments can be made for SH2 binding. For example, only three point mutations are sufficient to enhance the affinity of the FYN SH2 domain for the phosphorylated EGFR tyrosine residue 978 by a factor of 380, suggesting an evolutionary optimization for dynamics rather than binding strength (Kaneko et al. 2012a). In surface plasmon resonance-validated fluorescence popularization assays with ErbB peptides and SH2 domains, (Hause et al. 2012) found many weak high confidence interactions, as well as interactions with tyrosine residues previously not described as phosphorylated. In short, the structure of the phosphotyrosine-dependent protein-protein interaction network strongly suggests that the process of cellular decision making regarding growth and development is reflected in the interplay of phosphotyrosine-dependent protein-protein interactions with tyrosine phosphorylation and dephoshorylation.

4.3.2.3 The phosphotyrosine-dependent protein interactions reflect and connect many signaling pathways

In this study, we present a set of 292 tyrosine phosphorylation-dependent protein-protein interactions. Most of these are novel and many involve proteins that have not been studied intensely. Therefore, the data set is well-suited for hypothesis generation. We demonstrate how the findings provided here apply to the

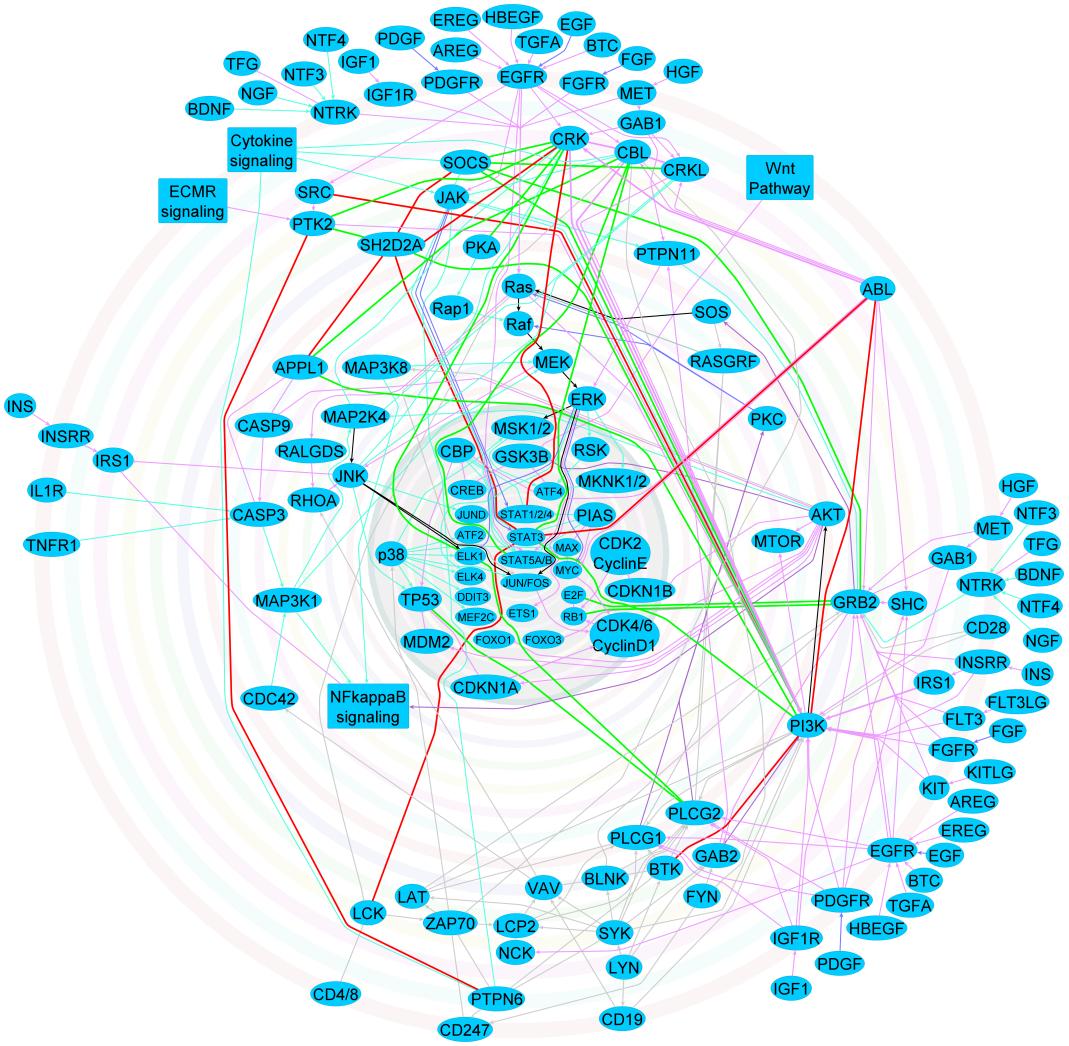


Figure 24: Established pathways provide biological context for phosphotyrosine-dependent interactions. Overview of selected interactions from the KEGG pathways related to phosphotyrosine signaling and phosphotyrosine-dependent protein-protein interactions found in this study. KEGG entities were mapped to Entrez GeneIDs using the KEGG annotations and reaggregated for greatest clarity. Protein components are shown as elliptical nodes, pathways are represented as squares. Edge directions were adopted from KEGG pathways where applicable. Edge were colored according to source. Edges originating from Cancer/Growth factor signaling KEGG pathways (hsa04012, hsa04910, hsa05200, hsa05210, hsa05211, hsa05212, hsa05213, hsa05214, hsa05218, hsa05219, hsa05220, hsa05222 and hsa05223) are purple, edges originating from MAPK, Toll-like receptor or Jak-STAT signaling KEGG pathways (hsa04010, hsa04620, hsa04630) are cyan, edges from immune-related KEGG pathways (hsa04660, hsa04662, hsa04664, hsa04666) are grey, edges occurring in more than one of these groups are black and edges that are not in these KEGG pathways are green. Thick edges represent phosphotyrosine-dependent protein-protein interaction found in this study. The components

pathways in the KEGG database, because this database is generally considered a rather conservative knowledgebase, making it a great point of reference. Figure 24 shows an aggregation of KEGG pathways related to functions associated with tyrosine phosphorylation, like growth factor receptor and immune signaling or different kinds of cancer. In addition, selected phosphorylation-dependent protein-protein interaction detected in this study are shown. Some of the added interactions are confirmatory, others suggest the existence of previously unappreciated regulatory loops or cross-pathway signal integration points.

4.3.2.3.1 The high number of KEGG pathway connections in the phosphotyrosine-dependent interaction network confirms its relevance

There are seven links in the network of phosphotyrosine protein-protein interactions that are reflected as direct links in the KEGG pathways database. These are the interactions between CRK and PTK2 (Akagi et al. 2002), CRK and PI3K (Akagi et al. 2002), IRS1 and PI3K (Felder et al. 1993; Pons et al. 1995; Mothe et al. 1997), CBL and CRK (Fukazawa et al. 1996), CBL and CRKL (Andoniou et al. 1996), CBL and PI3K (Fukazawa et al. 1996; Hunter et al. 1999), as well as between CRK and ABL (Ren et al. 1994). Most of these, specifically CRK-PI3K, CBL-CRK, CBL-CRKL, CBL-PI3K and CRK-ABL, are related to chronic myeloid leukemia. The interaction between IRS1 and PI3K is central to the insulin signaling pathway, as is the interaction between CBL and CRKL. The interaction between CRK and PTK2 is involved in the bacterial invasion of epithelial cells pathway.
Since the KEGG pathway interactions reflect information flow through biological systems, it is to be expected that some of the KEGG interactions, especially for pathways related to growth and development, immune functions and cancer, are

---

were arranged in a circular pattern maximizing the number of directed edges pointing inwards and interactions of similar KEGG pathway source were tried to place together. Where possible, protein components that are found predominantly in the nucleus were placed within the inner, slightly darker part of the figure. For greater clarity, some of the extracellular components and their membrane-bound receptors were split and occur more than once.

identical to phosphotyrosine-dependent protein interactions, as well as tyrosine phosphorylation and dephosphorylation events. Therefore, some overlap between the phosphotyrosine-dependent protein-protein interaction network and the KEGG pathways is to be expected and confirms the relevance and accuracy of the phosphotyrosine-dependent interaction network.

4.3.2.3.2 A direct link between PI3K and BTK provides additional feed-back possibilities in the Fcε receptor pathway

The interaction between PI3K and BTK is included in the Fcε receptor pathway (KEGG hsa04664) as an indirect interaction. PI3K function is thought to promote BTK membrane localization by creating PIP3, which is then bound by the BTK PH domain (Salim et al. 1996; Rameh et al. 1997). Membrane association has been shown to potentiate BTK activation (Li et al. 1997). The direct phosphotyrosine-dependent interaction strengthens the link between PI3K and BTK. Furthermore, it is possible that one of the two proteins works as a scaffold, allowing the other one to interact with its interaction partners. For example, BTK has been found to bind phosphatidylinositol-4-phosphate-5 kinases and account for its membrane localization (Carpenter 2004). Phosphatidylinositol-4-phosphate-5 kinases synthesize phosphatidylinositol-(4,5)-bisphosphate, the substrate for PI3K. Therefore, BTK can stimulate the output of PI3K by raising the local concentration of its substrate. Conversely, BTK is a substrate of SYK (Baba et al. 2001) and ABL (Bäckesjö et al. 2002). PI3K is phosphorylated by SYK in the Fcγand Fcεreceptor signaling pathways (KEGG hsa04664 & hsa04666) and was found as a phosphotyrosine-dependent interactor of ABL2. Thus, PI3K can also act as a scaffold for BTK phosphorylation.

4.3.2.3.3 Phosphotyrosine-dependent protein interactions connect and modify signaling pathways

More interesting than the overlap between KEGG pathways and the phosphotyrosine-dependent protein-protein interactions reported in this study in terms of scientific advancement are interactions that improve upon the current knowledge. Those interactions typically connect or ”short-circuit” established pathways, i.e. that connect two or more pathways that result similar responses

to different inputs or directly connect components of a single pathway not connected in the canonical view, respectively. The latter are especially interesting where they are phosphotyrosine-dependent, because they are likely to constitute optional or delayed feed-forward or feed-back mechanisms. The former are important, because they can provide valuable hints for augmenting and integrating pathway knowledge.

4.3.2.3.3.1 Phosphotyrosine-dependent interactions between components of the ErbB signaling pathway might explain the ability to appropriately react to different growth factors

Different biological signals going through the epidermal growth factor receptor family result in different biological responses (Graus-Porta et al. 1997; Hackel et al. 1999; Zwick et al. 1999). The canonical ErbB signaling pathway (KEGG hsa04012) splits up almost immediately downstream of the ErbB receptors. The different receptors, constituted by different combinations of ErbB monomers, affect specific downstream proteins leading to specific cellular responses. Our data suggest that the different branches of the ErbB signaling pathway are connected. The PI3 kinase, which signals through AKT1, and is downstream of the ErbB dimers binding neuregulins (NRG1-4), namely ErbB2-ErbB3, ErbB2-ErbB4 and ErbB4-ErbB4, has been found to interact in a phosphotyrosine-dependent manner with the tyrosine kinases SRC and Abl. These kinases are downstream targets of the EGF, TNFα and AR binding receptors, specifically ErbB1-ErbB1 and ErbB1-ErbB2. Similarly, ELK1, a transcription factor regulated through the Ras-Raf-ERK axis, has been shown to interact, in a phosphotyrosine-dependent manner, with CRK and PLCG2, both of which are downstream targets of the EGF, TNFα and AR binding receptors only. It is thus possible that they act antagonistically, sharpening the specificity of the response to different growth factors, or cooperatively, differentiating responses to BTC, HB-EGF and EREG, which can bind to almost all ErbB dimers, from the others.

4.3.2.3.3.2 The adaptor GRB2 has membrane-integral, cytoplasmic and nuclear interaction partners

A particularly interesting pair of links are the interactions of GRB2 with RB1 and E2F. RB1 and E2F are nuclear regulators of cell cycle progression (Dyson 1998). In healthy tissue, RB1 binds and inhibits E2F. Release of this

inhibition is a major hallmark of cancer progression (Hanahan & Weinberg 2000). GRB2 is an adapter molecule that is usually thought of as membrane-associated, immediately downstream of receptor tyrosine kinases. GRB2 has been appreciated as a binding partner of nuclear proteins only recently (Hino et al. 2011; Bisson et al. 2011), although (Verbeek et al. 1997) reported in 1997 already, that the relative amount of GRB2 in the nucleus is increased in breast cancer tissue. RB1 binds E2F directly (Helin et al. 1993). Therefore, the fact that GRB2 interacts with both, RB1 and E2F, suggests a more complex function than e.g. simple scaffolding or sequestering of interaction partners. It also emphasizes the importance of nuclear GRB2 and provides new leads for further study. Many of the GRB2 interactors found in this study were also nuclear. Because protein-protein interactions between proteins localized at different subcellular localizations are often suspected to be technical or biological false-positive, we validated several of these interactions by co-immunoprecipitation from mammalian cells. We further asked, where the interactions take place in the mammalian cell. We used protein complementation assays to establish that GRB2 can interacts in different subcellular localizations, depending on its interaction partners. Comparing the localization patterns of endogenous and over-expressed proteins to interaction signals, it seems that GRB2 meets its interaction partners in their respective localizations: the interaction between GRB2 and RB1 takes place in the nucleus, the interaction with C10orf81 is cytoplasmic and the interaction between GRB2 and TSPAN2 occurs at the membrane. Hence, the dynamics of GRB2 interactions seem to be regulated at several levels. Protein levels are regulated by modulation of expression and degradation, effective protein concentration depends on localization and tyrosine phosphorylation is constantly and potentially rapidly adjusted by a multitude of kinases and phosphatases.

4.3.2.3.3.3 The JAK/STAT signaling pathways is highly integrated indicating an early evolutionary origin

The JAK/STAT signaling pathway is another pathway for which phosphotyrosine binding is important. The pathway is triggered by cytokines, like interferons or interleukins, binding to transmembrane cytokine receptors. This activates Janus kinases (JAK1-3), which phosphorylate signal transducer and activator of transcription (STAT) proteins, as well as other cytoplasmic proteins,

like PI3K and PTPN11. The STAT proteins are transcription factors that dimerize upon phosphorylation and translocate to the nucleus, where they bind to DNA and regulate gene expression. The JAK/STAT pathway is down-regulated by Suppressor of Cytokine Signaling (SOCS) proteins through a negative feedback mechanism. Typically, the SOCS proteins are not expressed in resting cells. They are induced upon cytokine receptor binding. In fact, SOCS protein induction is usually mediated, at least partially, by STAT proteins (Starr & Hilton 1998). Once expressed, the SOCS proteins inhibit JAK/STAT signaling by competing with Janus kinase substrates or by marking Janus kinases for proteosomal degradation (Alexander & Hilton 2004). In this study, we found STAT5A interacting with CBL and STAT3 interacting with PIK3R3 and CBL, both of which are members of the KEGG JAK/STAT signaling pathway, but not in the same branch as the STATs. This points to the existence of mechanisms attuning the different branches. Interestingly, CBL overexpression has been shown to cause ubiquitination and degradation of STAT5 but not STAT3 in pUC-CAGGS-transfected NIH3T3 cells (Goh et al. 2002). Such a mechanism might have evolved to integrate JAK/STAT signaling in a larger phosphotyrosine signaling context. Most of the components of the JAK/STAT signaling pathway are members of other pathways as well. There is an especially intimate connection between JAK/STAT and insulin signaling. This is exemplified most strikingly by the cytokine Leptin, which acts through both, the JAK/STAT and the insulin pathway (Carvalheira et al. 2003). Fittingly, there is a cross-talk between JAK/STAT and IGF1 signaling that is modulated by SOCS proteins (Himpe & Kooijman 2009). STAT proteins have been found in *Dictyostelium discoideum*, a lower eukaryote, making them probably the evolutionarily oldest SH2 genes (Kawata et al. 1997), especially in light of the actin gene phylogeny (Drouin et al. 1995). Accordingly, the JAK/STAT signaling pathway might be the oldest signaling pathway utilizing tyrosine phosphorylation, acting as a foundation for the evolution of phosphotyrosine signaling. In this context, it is interesting to note that we found phosphotyrosine-dependent interactions between SOCS and the insulin pathway components PI3K, GRB2, CRK and CRKL. PI3K and GRB2 are components of the insulin signaling pathway and the JAK/STAT signaling pathway; all four are connections between different branches of the respective pathways, again providing a network structure suit-

able for the integration of different input signals affecting the same output. Canonically, the ubiquitination function of SOCS is thought to operate through TCEB1/2 (Zhang et al. 1999). The interactions of SOCS with CRK and CRKL provide hints towards alternative regulatory mechanisms for the degradation of JAK or IRS1/2. CRK and CRKL are often associated with the E3 ubiquitinase CBL that might contribute to JAK down-regulation. Conversely, TCEB1/2 could be considered as additional ubiquitin ligases for CRK/CRKL targets, as Rui et al. (2002) suggested for IRS1 and IRS2. IRS1 and IRS2 degradation has been shown to be mediated by several ubiquitin ligases, like CBL (Nakao et al. 2009) and CUL7 (Xu et al. 2008). In this context, the interaction of STAT4 and CRK is similar to the interaction of STAT3 and CBL. Both are interactions between STATs and ubiquitination-related proteins and might reflect synergistic or antagonistic effects of multiple cytokines.

The interactions of STAT3 with LCK, Abl and SH2D2A seem to be related to STAT3 function in T cells, where it has been shown to be responsible for IL6-induced proliferation (Takeda et al. 1998). Here, STAT3 (but not STAT5) is phosphorylated upon anti-CD3 stimulation (Gerwien et al. 1999), supporting an immunological context for the interaction with SH2D2A. This phosphorylation can be blocked by the Src family and Abl-specific inhibitor PP1 (Gerwien et al. 1999). Therefore, the interactions of STAT3 with LCK and Abl probably represent kinase-substrate relationships. Indeed, the interaction of STAT3 with Abl has been described as a minor kinase-substrate relationship (Ilaria & Van Etten 1996). The respective major Abl substrate STAT5A has been shown to be important for the transformation of K562 leukemia cells (de Groot et al. 1999). Interestingly, STAT3 had by far the most phosphorylation-dependent interactions in our study. STAT3 interacted with 46 other proteins, while STAT1, STAT2, STAT4 and STAT5A had only 1, 1, 2 and 2 interactions, respectively. Given yeast two-hybrids tendency to favor specific genes, or even clones, for non-apparent reasons, this might be a coincidence. On the other hand, there is considerable evidence pointing toward a special role of STAT3. Genetically, STAT genes are organized in 3 tandem clusters, STAT1-STAT4, STAT2-STAT6 and STAT3-STAT5 with a relatively recent split of STAT5 into STAT5A and STAT5B (Copeland et al. 1995). STAT3-STAT5 seems to be the oldest cluster (Copeland et al. 1995; Miyoshi et al. 2001). Apparently, STAT3 and STAT5 are

also the only STAT genes that have transforming potential and are activated by a larger number of cytokines and more pleiotropic in their function than the other STAT genes (Kisseleva et al. 2002). STAT3 seems to play an important role in a higher number of different biological processes than other STAT genes (Hirano et al. 2000; Levy & Lee 2002). Furthermore, STAT3 knockout mice die at early during embryogenesis (E6.5-E7.5) (Takeda et al. 1997), while the other STAT gene knockouts have much more moderate phenotypes (O'Shea 1997). Finally, STAT3 has been proposed to function not only as a transcription factor but also as an adapter, e.g. linking PI3K and the interferon 1 receptor (Pfeffer et al. 1997).
On the whole, the KEGG pathway overlay supports the conclusions drawn from the network statistics. The core nodes STAT3, SOCS4, GRB2, PIK3R3, CBL and CRK are all important members of the JAK/STAT signaling pathway and take central positions in the aggregated KEGG pathway network. This emphasizes the importance of the core nodes and suggest that the JAK/STAT signaling pathway is the most strongly inter-connected pathway in the context of phosphotyrosine signaling.

4.3.2.3.3.4 The adaptor SH2D2A appears to be an under-appreciated signaling hub

Another core gene that links several KEGG pathways by phosphotyrosine-dependent interactions is SH2D2A. The highest expression levels for this gene have been measured in activated T and NK cells, as well as certain types of endothelial and epithelial cells (Berge et al. 2012). Accordingly, a nucleotide polymorphism in the promoter region of the SH2D2A gene has been linked to reduced protein levels and susceptibility to multiple sclerosis (Dai et al. 2001) and SH2D2A knock-out mice show severe disturbances in the fine structures of the basal membrane and intercellular epithelial spaces (Kolltveit et al. 2010). Lapinski et al. (2009) argue that one would expect a more severe phenotype, but the fact that SH2D2A, SH2D4A and SH2D4B have overlapping functions, is masking some knock-out effects. Initially, SH2D2A was described as a binder of LCK (Choi et al. 1999), ITK and TXK (Rajagopal et al. 1999), important in the activation of T cells. However, its exact function is still debated, owed in no small part to its connection to cAMP, which is intimitely entangled in T cell signaling. The SH2D2A promoter contains a cAMP response element

that is critically important for transcriptional regulation in T cells (Dai et al. 2004), as does the T cell receptor itself (Lee et al. 1992). SH2D2A protein concentration is regulated on two levels: cAMP-mediated signals are sufficient to induce transcription of the SH2D2A gene, whereas TCR/CD3-dependent signals are necessary for the translation of the transcript (Kolltveit et al. 2008). The SH2D2A protein is widely agreed to regulate LCK kinase activity, but several models have been submitted explaining the mechanism of action. Marti et al. (2006) showed that SH2D2A binds the SH2 and the SH3 domain of LCK, releasing intramolecular inhibition. This was proposed to follow LCK phosphorylation by activated T cell receptors (Lapinski et al. 2009). Since SH2D2A and LCK are usually found in a 1:1 ratio, others speculated that SH2D2A acts as a competitive LCK inhibitor, effectively sequestering it from other potential targets (Granum et al. 2008). Both models have merit and are not entirely exclusive. SH2D2A might actually have an inhibiting as well as an activating effect on LCK. Similarly, PTPRC, the phosphatase responsible for removing the activating phospho group on LCK tyrosine residue 394, also removes the inhibitory phospho group on LCK tyrosine residue 505 (D'Oro & Ashwell 1999). The inhibitory site LCK-Y505 is phosphorylated by CSK, down-regulating TCR signaling and IL-2 production (Vang et al. 2001). CSK, in turn, is activated by protein kinase A (PKA) through phosphorylation of CSK-S364. PKA is regulated by cAMP and colocalizes with the TCR-CD3 complex in primary peripheral T cells (Skålhegg et al. 1994). It also effects CREB phosphorylation. In normal peripheral T cells and thymocytes CREB phosphorylation can be mediated by at least three distinct pathways: a PKA-dependent pathway, a Ca^{2+}/calmodulin-dependent pathway, and a PKC-dependent pathway (Muthusamy & Leiden 1998). Thus, the different modes of cAMP action circle back and converge on the CRE-mediated regulation of gene expression, contributing to the difficulties in breaking down the exact function of SH2D2A in T cells.

However, both of models of LCK regulation by SH2D2A fail to satisfactorily explain the function of the SH2D2A SH2 domain. This domain might bestow an adapter function upon SH2D2A. In this study, protein kinase A (PKA) interacted with CRK in a phosphotyrosine-dependent manner. As a consequence, the phosphotyrosine-dependent interaction of SH2D2A and CRK may provide

a feedback mechanism and a connection of T cell receptor signaling and MAP kinase signaling. Furthermore, in PC12 cells, cAMP has been demonstrated to activate ELK1, another phosphotyrosine-dependent interactor of CRK found in this study (Vossler et al. 1997). Moreover, binding SH2D2A could allow LCK to further phosphorylate other phosphotyrosine proteins. This behavior is in good accordance with the progressive phosphorylation model (Zhao et al. 2013) and findings reported in context with VEGFR signaling. SH2D2A attaches to the VEGF receptor and becomes phosphorylated (Matsumoto et al. 2005). It further binds SRC and other Src family kinases to affect cytoskeletal rearrangements (Matsumoto et al. 2005). FRS2, one of the main targets of the VEGF receptors, acts as an adapter for SRC, which binds it through its SH3 domain (Meakin et al. 1999). Meakin et al. (1999) proposed that competition for VEGF receptor binding between FRS2 and SHC1 is a key mechanism for the differentiation of proliferation and differentiation signals. Possibly, SH2D2A fulfills a similar function, either in VEGF or T cell receptor signaling. In our study, SH2D2A interacted with several bona fide signaling hubs, each of which plays important roles in several pathways. Notably, Berg & Ostergaard (1997) showed a direct association between PTK2 and the SH2 domain of the SH2D2A interacting kinase LCK. Therefore, if connecting PTK2 and LCK is a biologically relevant function of SH2D2A, it is probably regulated by different kinases than the direct interaction. It seems more plausible that SH2D2A connects both PTK2 and LCK to other proteins. For example, PTK2 becomes phosphorylated upon TCR activation, even in $LCK^{-/-}$ cells (Berg & Ostergaard 1997). Correspondingly, in endothelial cells, SH2D2A binds VEGFR2 on phosphotyrosine 951 and mutating this tyrosine residue to phenylalanine blocks VEGFA-induced actin stress fibers and migration, but not mitogenesis (Matsumoto et al. 2005). In contrast, in TCR signaling, the PTK2-related kinase ZAP70 is a substrate of LCK and signals through p38. In endothelial cells, VEGFR2-pY1214 is required for Rho/Rac-independent activation of CDC42 and p38 (Lamalice et al. 2004). In porcine aortic endothelial cells expressing human VEGFR2, SHB has been shown to bind VEGFR2 phopsphotyrosine 1175 after being phosphorylated in a manner sensitive to the Src family kinase inhibitor PP2 (Holmqvist et al. 2004). After binding to VEGFR2, SHB activates PTK2 and PI3K, both of which interacted with SH2D2A in phosphotyrosine-dependent

manner in this study. Interestingly, the association between SH2D2A and PI3K has been shown to be independent of VEGF treatment in human endothelial cells (Wu et al. 2000). The most obvious function of SH2D2A in vivo is the induction of superantigen-induced T cell death, exemplified by a raised susceptibility to lupus-like autoimmune diseases (Marti et al. 2005). T cells have two apoptosis pathways: one operating through FAS-FADD, CASP8-CASP3 and a "mitochondrial" one operating through CytC-CASP9-CASP3 (Hildeman et al. 2002b). SH2D2A seems to play a role in the induction of this intrinsic FAS-independent mechanism for programmed cell death, which is most probably mediated by BCL2L11 (Hildeman et al. 2002a). SH2D2A exerts its effects through the transcriptional regulation of interleukin 2 (IL-2) and other genes, but does not appear to modulate other regulators of apoptosis like NF-kappaB, NFAT or AP-1 (Marti et al. 2001). The interaction of SH2D2A and APPL1 seems immediately relevant in this context, since disruption of APPL1 signaling through CASP9/CASP3 is known to contribute to colorectal cancer (compare KEGG pathway hsa05210). The IL-2 promoter contains many binding motifs, e.g. for Jun, Fos, AP1, ELK1 and NFAT (Serfling et al. 1995). It has been speculated that other factors bind the IL-2 promoter indirectly (Serfling et al. 1995). In our study, SH2D2A interacted with many proven or putative transcriptional regulators, like ARID5A (Patsialou et al. 2005), LRRFIP1 (Reed et al. 1998), NAA16 (inferred from structural similarity), RAD54B (Wesoly et al. 2006), ZHX3 (Yamada et al. 2003) or ZSCAN1 (inferred from electronic annotation). This casts SH2D2A as a nuclear adapter for transcription factors and regulators. This role fits the findings of (Drappa et al. 2003) who analyze differences in mRNA expression upon SH2D2A knock-out in C57BL/6 mice and find many transcription factors and other genes related DNA damage response and apoptosis down-regulated.

In short, the SH2D2A gene encodes an adapter protein with important functions in the cytoplasm and in the nucleus. A similar characterization fits several of SH2D2As interaction partners: APPL1, PIK3R3, STAT3 and GRB2. The last interaction was not found in this study, but has been described to be stimulated by PDGF treatment in human embryonic lung cells (Park et al. 2001) and shown to be phosphotyrosine-dependent in yeast two-hybrid experiments elsewhere (Park & Yun 2001). The nuclear import of SH2D2A is dependent on its SH2

domain and a phosphorylation on the tyrosine residue 805 in the AAA ATPase VCP, although a direct binding has not been shown (Marti & King 2005). In this study, VCP interacted with GRB2 in a phosphorylation-dependent manner. This suggests a connection between the nuclear import mechanisms of SH2D2A and GRB2. Interestingly, VCP is necessary for ionizing radiation-induced 53BP1 recruitment to DNA double strand breaks (Acs et al. 2011; Meerang et al. 2011). Thus, it is probable that the connection between SH2D2A and DNA damage response is connected to TP53 function.

In conclusion, our results suggest that SH2D2A plays a much more important role as a signaling hub than is currently appreciated.

4.3.3 Conclusion

In this study, we present a comprehensive screen for phosphotyrosine-dependent interactions among human proteins. Using a yeast two-hybrid system that employs non-receptor tyrosine kinases to allow tyrosine phosphorylation. We screened a total of 159 clones, representing 121 genes containing known phosphotyrosine-recognizing domains, against a proteome-scale prey matrix. We obtained 292 phosphotyrosine-dependent interactions involving 86 phosphotyrosine-recognizing domain-containing genes. The high quality of the collected interaction data was demonstrated in high throughput co-immunoprecipitation experiments from mammalian cell culture. The interactions were compared to the literature and are mostly novel. This emphasizes the strong utility of a yeast two-hybrid screen complementing the prior knowledge about phosphotyrosine signaling, which comprised mostly peptide array- and affinity purification-coupled mass spectrometry-based experiments. The network of phosphotyrosine-dependent interactions is extremely dense and genes associated with known phosphotyrosine signaling-related processes, pathways and disease are over-represented among the interaction partners . Consistent with other protein-protein interactions, the phosphotyrosine-dependent interactions form a scale-free network, albeit one with an unusually steep degree distribution, possibly indicating a unique property of post translational modification-dependent protein interaction networks. The closely connected single-module structure of the phosphotyrosine-dependent protein interaction network contrast with the possibility of several discrete modules responsible for different biological functions. This structure suggests that the high number of growth- and development-related signals relayed by a multitude of growth factor/receptor tyrosine kinase binding events are integrated by the phosphotyrosine-dependent interaction network. This picture is in accordance with reports finding highly overlapping tyrosine kinase substrate pairings or that many SH2 domains usually recognize more than one binding site on the same protein. In this study, we show that, even in the artificial yeast environment, there is a specificity of interactions for the assayed kinases, that is not explained by one of the interacting proteins alone, lending further credence to this view. The phosphotyrosine-dependent interaction partners of SH2 domain-containing proteins are enriched for the

respective reported linear peptide binding motifs, but the motifs are unable to explain all of the interactions. What is more, we closely examined selected phosphotyrosine-dependent GRB2 interactions with conflicting subcellular localizations using protein complentation assays in living mammalian cells and found that GRB2 displayed a high motility allowing it to meet its interaction partners at the plasma membrane, in the cytoplasm or in the nucleus. Finally, we analyzed the phosphotyrosine-dependent interactions in the context of the conservative KEGG pathway knowledgebase. In this analysis, we find not only confirmatory overlaps and complementary cross-pathway interactions, we also identify SH2D2A as an under-appreciated signaling hub. To sum up, we provide a comprehensive, high-quality set of phosphotyrosine-dependent protein-protein interaction that lays the foundation for a plethora of medically relevant analyses.

Bibliography

Acs, K., Luijsterburg, M.S., Ackermann, L., Salomons, F.A., Hoppe, T., Dantuma, N.P. The AAA-ATPase VCP/p97 promotes 53BP1 recruitment by removing L3MBTL1 from DNA double-strand breaks. *Nature Structural & Molecular Biology*, 18(12):1345–1350, 2011.

Acuto, O., Di Bartolo, V., Michel, F. Tailoring T-cell receptor signals by proximal negative feedback mechanisms. *Nature Reviews. Immunology*, 8(9):699–712, 2008.

Aebersold, R., Goodlett, D.R. Mass spectrometry in proteomics. *Chemical Reviews*, 101(2):269–295, 2001.

Aebersold, R., Mann, M. Mass spectrometry-based proteomics. *Nature*, 422(6928):198–207, 2003.

Akagi, T., Murata, K., Shishido, T., Hanafusa, H. v-Crk activates the phosphoinositide 3-kinase/AKT pathway by utilizing focal adhesion kinase and H-Ras. *Molecular and Cellular Biology*, 22(20):7015–7023, 2002.

Albert, R., Barabási, A.L. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

Alexander, W.S., Hilton, D.J. The role of suppressors of cytokine signaling (SOCS) proteins in regulation of the immune response. *Annual Review of Immunology*, 22:503–529, 2004.

Alonso, A., Sasin, J., Bottini, N., Friedberg, I., Friedberg, I., Osterman, A., Godzik, A., Hunter, T., Dixon, J., Mustelin, T. Protein tyrosine phosphatases in the human genome. *Cell*, 117(6):699–711, 2004.

Andersen, J.N., Mortensen, O.H., Peters, G.H., Drake, P.G., Iversen, L.F., Olsen, O.H., Jansen, P.G., Andersen, H.S., Tonks, N.K., Møller, N.P. Structural and evolutionary relationships among protein tyrosine phosphatase domains. *Molecular and Cellular Biology*, 21(21):7117–7136, 2001.

Andoniou, C.E., Thien, C.B., Langdon, W.Y. The two major sites of cbl tyrosine phosphorylation in abl-transformed cells select the crkL SH2 domain. *Oncogene*, 12(9):1981–1989, 1996.

Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J.K., Lonai, P. Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. *Proceedings of the National Academy of Sciences of the United States of America*, 95(9):5082–5087, 1998.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25(1):25–29, 2000.

Baba, Y., Hashimoto, S., Matsushita, M., Watanabe, D., Kishimoto, T., Kurosaki, T., Tsukada, S. BLNK mediates Syk-dependent Btk activation. *Proceedings of the National Academy of Sciences of the United States of America*, 98(5):2582–2586, 2001.

Bader, G.D., Hogue, C.W.V. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*, 4:2, 2003.

Bae, J.H., Lew, E.D., Yuzawa, S., Tomé, F., Lax, I., Schlessinger, J. The selectivity of receptor tyrosine kinase signaling is controlled by a secondary SH2 domain binding site. *Cell*, 138(3):514–524, 2009.

Bae, Y.S., Oh, H., Rhee, S.G., Yoo, Y.D. Regulation of reactive oxygen species generation in cell signaling. *Molecules and Cells*, 32(6):491–509, 2011.

Bagowski, C.P., Bruins, W., Te Velthuis, A.J.W. The nature of protein domain evolution: shaping the interaction network. *Current Genomics*, 11(5):368–376, 2010.

Bandyopadhyay, S., Chiang, C.y., Srivastava, J., Gersten, M., White, S., Bell, R., Kurschner, C., Martin, C.H., Smoot, M., Sahasrabudhe, S., Barber, D.L., Chanda, S.K., Ideker, T. A human MAP kinase interactome. *Nature Methods*, 7(10):801–805, 2010.

Barabási, A.L., Gulbahce, N., Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature Reviews. Genetics*, 12(1):56–68, 2011.

Barabási, A.L., Oltvai, Z.N. Network biology: understanding the cell's functional organization. *Nature Reviews. Genetics*, 5(2):101–113, 2004.

Barilá, D., Superti-Furga, G. An intramolecular SH3-domain interaction regulates c-Abl activity. *Nature Genetics*, 18(3):280–282, 1998.

Barrios-Rodiles, M., Brown, K.R., Ozdamar, B., Bose, R., Liu, Z., Donovan, R.S., Shinjo, F., Liu, Y., Dembowy, J., Taylor, I.W., Luga, V., Przulj, N., Robinson, M., Suzuki, H., Hayashizaki, Y., Jurisica, I., Wrana, J.L. High-throughput mapping of a dynamic signaling network in mammalian cells. *Science (New York, N.Y.)*, 307(5715):1621–1625, 2005.

Batada, N.N., Hurst, L.D., Tyers, M. Evolutionary and physiological importance of hub proteins. *PLoS computational biology*, 2(7):e88, 2006.

Bauer, A., Kuster, B. Affinity purification-mass spectrometry. Powerful tools for the characterization of protein complexes. *European journal of biochemistry / FEBS*, 270(4):570–578, 2003.

Beausoleil, S.A., Jedrychowski, M., Schwartz, D., Elias, J.E., Villén, J., Li, J., Cohn, M.A., Cantley, L.C., Gygi, S.P. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proceedings of the National Academy of Sciences of the United States of America*, 101(33):12130–12135, 2004.

Bechtel, S., Rosenfelder, H., Duda, A., Schmidt, C.P., Ernst, U., Wellenreuther, R., Mehrle, A., Schuster, C., Bahr, A., Blöcker, H., Heubner, D., Hoerlein, A., Michel, G., Wedler, H., Köhrer, K., Ottenwälder, B., Poustka, A., Wiemann,

S., Schupp, I. The full-ORF clone resource of the German cDNA Consortium. *BMC genomics*, 8:399, 2007.

Beebe, K.D., Wang, P., Arabaci, G., Pei, D. Determination of the binding specificity of the SH2 domains of protein tyrosine phosphatase SHP-1 through the screening of a combinatorial phosphotyrosyl peptide library. *Biochemistry*, 39(43):13251–13260, 2000.

Benes, C.H., Wu, N., Elia, A.E.H., Dharia, T., Cantley, L.C., Soltoff, S.P. The C2 domain of PKCdelta is a phosphotyrosine binding domain. *Cell*, 121(2):271–280, 2005.

Benjamini, Y., Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300, 1995.

Berg, N.N., Ostergaard, H.L. T cell receptor engagement induces tyrosine phosphorylation of FAK and Pyk2 and their association with Lck. *Journal of Immunology (Baltimore, Md.: 1950)*, 159(4):1753–1757, 1997.

Berge, T., Grønningsæter, I.H.B., Lorvik, K.B., Abrahamsen, G., Granum, S., Sundvold-Gjerstad, V., Corthay, A., Bogen, B., Spurkland, A. SH2D2A modulates T cell mediated protection to a B cell derived tumor in transgenic mice. *PloS One*, 7(10):e48239, 2012.

Bhalla, U.S., Iyengar, R. Emergent properties of networks of biological signaling pathways. *Science (New York, N.Y.)*, 283(5400):381–387, 1999.

Bilsland, J.G., Wheeldon, A., Mead, A., Znamenskiy, P., Almond, S., Waters, K.A., Thakur, M., Beaumont, V., Bonnert, T.P., Heavens, R., Whiting, P., McAllister, G., Munoz-Sanjuan, I. Behavioral and neurochemical alterations in mice deficient in anaplastic lymphoma kinase suggest therapeutic potential for psychiatric indications. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 33(3):685–700, 2008.

Bisson, N., James, D.A., Ivosev, G., Tate, S.A., Bonner, R., Taylor, L., Pawson, T. Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nature Biotechnology*, 29(7):653–658, 2011.

Bladt, F., Riethmacher, D., Isenmann, S., Aguzzi, A., Birchmeier, C. Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud. *Nature*, 376(6543):768–771, 1995.

Blagoev, B., Kratchmarova, I., Ong, S.E., Nielsen, M., Foster, L.J., Mann, M. A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nature Biotechnology*, 21(3):315–318, 2003.

Blagoev, B., Ong, S.E., Kratchmarova, I., Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nature Biotechnology*, 22(9):1139–1145, 2004.

Blaikie, P., Immanuel, D., Wu, J., Li, N., Yajnik, V., Margolis, B. A region in Shc distinct from the SH2 domain can bind tyrosine-phosphorylated growth factor receptors. *The Journal of Biological Chemistry*, 269(51):32031–32034, 1994.

Blomberg, N., Baraldi, E., Nilges, M., Saraste, M. The PH superfold: a structural scaffold for multiple functions. *Trends in Biochemical Sciences*, 24(11):441–445, 1999.

Blume-Jensen, P., Hunter, T. Oncogenic kinase signalling. *Nature*, 411(6835):355–365, 2001.

Borowiec, M., Liew, C.W., Thompson, R., Boonyasrisawat, W., Hu, J., Mlynarski, W.M., El Khattabi, I., Kim, S.H., Marselli, L., Rich, S.S., Krolewski, A.S., Bonner-Weir, S., Sharma, A., Sale, M., Mychaleckyj, J.C., Kulkarni, R.N., Doria, A. Mutations at the BLK locus linked to maturity onset diabetes of the young and beta-cell dysfunction. *Proceedings of the National Academy of Sciences of the United States of America*, 106(34):14460–14465, 2009.

Boschelli, F., Uptain, S.M., Lightbody, J.J. The lethality of p60v-src in Saccharomyces cerevisiae and the activation of p34CDC28 kinase are dependent on the integrity of the SH2 domain. *Journal of Cell Science*, 105 (Pt 2):519–528, 1993.

Brand, M., Ranish, J.A., Kummer, N.T., Hamilton, J., Igarashi, K., Francastel, C., Chi, T.H., Crabtree, G.R., Aebersold, R., Groudine, M. Dynamic changes

in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nature Structural & Molecular Biology*, 11(1):73–80, 2004.

Braun, P., Tasan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., Yu, H., Sahalie, J.M., Murray, R.R., Roncari, L., de Smet, A.S., Venkatesan, K., Rual, J.F., Vandenhaute, J., Cusick, M.E., Pawson, T., Hill, D.E., Tavernier, J., Wrana, J.L., Roth, F.P., Vidal, M. An experimentally derived confidence score for binary protein-protein interactions. *Nature Methods*, 6(1):91–97, 2009.

Brena, R.M., Morrison, C., Liyanarachchi, S., Jarjoura, D., Davuluri, R.V., Otterson, G.A., Reisman, D., Glaros, S., Rush, L.J., Plass, C. Aberrant DNA methylation of OLIG1, a novel prognostic factor in non-small cell lung cancer. *PLoS medicine*, 4(3):e108, 2007.

Brown, M.T., Cooper, J.A. Regulation, substrates and functions of src. *Biochimica Et Biophysica Acta*, 1287(2-3):121–149, 1996.

Brugge, J.S., Jarosik, G., Andersen, J., Queral-Lustig, A., Fedor-Chaiken, M., Broach, J.R. Expression of Rous sarcoma virus transforming protein pp60v-src in Saccharomyces cerevisiae cells. *Molecular and Cellular Biology*, 7(6):2180–2187, 1987.

Bu, D., Zhao, Y., Cai, L., Xue, H., Zhu, X., Lu, H., Zhang, J., Sun, S., Ling, L., Zhang, N., Li, G., Chen, R. Topological structure analysis of the protein-protein interaction network in budding yeast. *Nucleic Acids Research*, 31(9):2443–2450, 2003.

Buggy, J.J., Elias, L. Bruton tyrosine kinase (BTK) and its role in B-cell malignancy. *International Reviews of Immunology*, 31(2):119–132, 2012.

Bunnell, S.C., Diehn, M., Yaffe, M.B., Findell, P.R., Cantley, L.C., Berg, L.J. Biochemical interactions integrating Itk with the T cell receptor-initiated signaling cascade. *The Journal of Biological Chemistry*, 275(3):2219–2230, 2000.

Butt, T.R., Sternberg, E.J., Gorman, J.A., Clark, P., Hamer, D., Rosenberg, M., Crooke, S.T. Copper metallothionein of yeast, structure of the gene, and

regulation of expression. *Proceedings of the National Academy of Sciences of the United States of America*, 81(11):3332–3336, 1984.

Bäckesjö, C.M., Vargas, L., Superti-Furga, G., Smith, C.I. Phosphorylation of Bruton's tyrosine kinase by c-Abl. *Biochemical and Biophysical Research Communications*, 299(3):510–515, 2002.

Campbell, I.D., Baron, M. The structure and function of protein modules. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 332(1263):165–170, 1991.

Cao, H., Courchesne, W.E., Mastick, C.C. A phosphotyrosine-dependent protein interaction screen reveals a role for phosphorylation of caveolin-1 on tyrosine 14: recruitment of C-terminal Src kinase. *The Journal of Biological Chemistry*, 277(11):8771–8774, 2002.

Carpenter, C.L. Btk-dependent regulation of phosphoinositide synthesis. *Biochemical Society Transactions*, 32(Pt 2):326–329, 2004.

Carpenter, C.L., Cantley, L.C. Phosphoinositide kinases. *Current Opinion in Cell Biology*, 8(2):153–158, 1996.

Carvalheira, J.B.C., Ribeiro, E.B., Folli, F., Velloso, L.A., Saad, M.J.A. Interaction between leptin and insulin signaling pathways differentially affects JAK-STAT and PI 3-kinase-mediated signaling in rat liver. *Biological Chemistry*, 384(1):151–159, 2003.

Chakrabarti, P., Janin, J. Dissecting protein-protein recognition sites. *Proteins*, 47(3):334–343, 2002.

Chan, B., Lanyi, A., Song, H.K., Griesbach, J., Simarro-Grande, M., Poy, F., Howie, D., Sumegi, J., Terhorst, C., Eck, M.J. SAP couples Fyn to SLAM immune receptors. *Nature Cell Biology*, 5(2):155–160, 2003.

Chao, M.V. Growth factor signaling: where is the specificity? *Cell*, 68(6):995–997, 1992.

Charbonnier, S., Gallego, O., Gavin, A.C. The social network of a cell: recent advances in interactome mapping. *Biotechnology Annual Review*, 14:1–28, 2008.

Chen, J., Nachabah, A., Scherer, C., Ganju, P., Reith, A., Bronson, R., Ruley, H.E. Germ-line inactivation of the murine Eck receptor tyrosine kinase by gene trap retroviral insertion. *Oncogene*, 12(5):979–988, 1996.

Cheng, A.M., Rowley, B., Pao, W., Hayday, A., Bolen, J.B., Pawson, T. Syk tyrosine kinase required for mouse viability and B-cell development. *Nature*, 378(6554):303–306, 1995.

Cheng, K.C.C., Klancer, R., Singson, A., Seydoux, G. Regulation of MBK-2/DYRK by CDK-1 and the pseudophosphatases EGG-4 and EGG-5 during the oocyte-to-embryo transition. *Cell*, 139(3):560–572, 2009.

Choi, Y.B., Kim, C.K., Yun, Y. Lad, an adapter protein interacting with the SH2 domain of p56lck, is required for T cell activation. *Journal of Immunology (Baltimore, Md.: 1950)*, 163(10):5242–5249, 1999.

Chothia, C., Lesk, A.M. The relation between the divergence of sequence and structure in proteins. *The EMBO journal*, 5(4):823–826, 1986.

Choudhuri, S. Gene regulation and molecular toxicology. *Toxicology Mechanisms and Methods*, 15(1):1–23, 2004.

Christofk, H.R., Vander Heiden, M.G., Wu, N., Asara, J.M., Cantley, L.C. Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature*, 452(7184):181–186, 2008.

Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D., Ideker, T. Network-based classification of breast cancer metastasis. *Molecular Systems Biology*, 3:140, 2007.

Citri, A., Skaria, K.B., Yarden, Y. The deaf and the dumb: the biology of ErbB-2 and ErbB-3. *Experimental Cell Research*, 284(1):54–65, 2003.

Clements, J.L., Yang, B., Ross-Barta, S.E., Eliason, S.L., Hrstka, R.F., Williamson, R.A., Koretzky, G.A. Requirement for the leukocyte-specific adapter protein SLP-76 for normal T cell development. *Science (New York, N.Y.)*, 281(5375):416–419, 1998.

Cohen, R., Havlin, S. Scale-free networks are ultrasmall. *Physical Review Letters*, 90(5):058701, 2003.

Collins, S.R., Kemmeren, P., Zhao, X.C., Greenblatt, J.F., Spencer, F., Holstege, F.C.P., Weissman, J.S., Krogan, N.J. Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. *Molecular & cellular proteomics: MCP*, 6(3):439–450, 2007.

Colvin, J.S., Bohne, B.A., Harding, G.W., McEwen, D.G., Ornitz, D.M. Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3. *Nature Genetics*, 12(4):390–397, 1996.

Copeland, N.G., Gilbert, D.J., Schindler, C., Zhong, Z., Wen, Z., Darnell, J.E., Mui, A.L., Miyajima, A., Quelle, F.W., Ihle, J.N. Distribution of the mammalian Stat gene family in mouse chromosomes. *Genomics*, 29(1):225–228, 1995.

Croce, C.M. Oncogenes and cancer. *The New England Journal of Medicine*, 358(5):502–511, 2008.

Croker, B.A., Kiu, H., Nicholson, S.E. SOCS regulation of the JAK/STAT signalling pathway. *Seminars in Cell & Developmental Biology*, 19(4):414–422, 2008.

Cuccato, G., Della Gatta, G., di Bernardo, D. Systems and Synthetic biology: tackling genetic networks and complex diseases. *Heredity*, 102(6):527–532, 2009.

Cusick, M.E., Klitgord, N., Vidal, M., Hill, D.E. Interactome: gateway into systems biology. *Human Molecular Genetics*, 14 Spec No. 2:R171–181, 2005.

Cusick, M.E., Yu, H., Smolyar, A., Venkatesan, K., Carvunis, A.R., Simonis, N., Rual, J.F., Borick, H., Braun, P., Dreze, M., Vandenhaute, J., Galli, M., Yazaki, J., Hill, D.E., Ecker, J.R., Roth, F.P., Vidal, M. Literature-curated protein interaction datasets. *Nature Methods*, 6(1):39–46, 2009.

Dai, K., Liao, S., Zhang, J., Zhang, X., Tu, X. Solution structure of tensin2 SH2 domain and its phosphotyrosine-independent interaction with DLC-1. *PloS One*, 6(7):e21965, 2011.

Dai, K.Z., Harbo, H.F., Celius, E.G., Oturai, A., Sørensen, P.S., Ryder, L.P., Datta, P., Svejgaard, A., Hillert, J., Fredrikson, S., Sandberg-Wollheim, M.,

Laaksonen, M., Myhr, K.M., Nyland, H., Vartdal, F., Spurkland, A. The T cell regulator gene SH2D2A contributes to the genetic susceptibility of multiple sclerosis. *Genes and Immunity*, 2(5):263–268, 2001.

Dai, K.Z., Johansen, F.E., Kolltveit, K.M., Aasheim, H.C., Dembic, Z., Vartdal, F., Spurkland, A. Transcriptional activation of the SH2D2A gene is dependent on a cyclic adenosine 5'-monophosphate-responsive element in the proximal SH2D2A promoter. *Journal of Immunology (Baltimore, Md.: 1950)*, 172(10):6144–6151, 2004.

Dai, X.M., Ryan, G.R., Hapel, A.J., Dominguez, M.G., Russell, R.G., Kapp, S., Sylvestre, V., Stanley, E.R. Targeted disruption of the mouse colony-stimulating factor 1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency, increased primitive progenitor cell frequencies, and reproductive defects. *Blood*, 99(1):111–120, 2002.

Darnell, J.E. Phosphotyrosine signaling and the single cell:metazoan boundary. *Proceedings of the National Academy of Sciences of the United States of America*, 94(22):11767–11769, 1997.

Deane, J.A., Fruman, D.A. Phosphoinositide 3-kinase: diverse roles in immune cell activation. *Annual Review of Immunology*, 22:563–598, 2004.

DeChiara, T.M., Bowen, D.C., Valenzuela, D.M., Simmons, M.V., Poueymirou, W.T., Thomas, S., Kinetz, E., Compton, D.L., Rojas, E., Park, J.S., Smith, C., DiStefano, P.S., Glass, D.J., Burden, S.J., Yancopoulos, G.D. The receptor tyrosine kinase MuSK is required for neuromuscular junction formation in vivo. *Cell*, 85(4):501–512, 1996.

Deepa, S.S., Zhou, L., Ryu, J., Wang, C., Mao, X., Li, C., Zhang, N., Musi, N., DeFronzo, R.A., Liu, F., Dong, L.Q. APPL1 mediates adiponectin-induced LKB1 cytosolic localization through the PP2A-PKCzeta signaling pathway. *Molecular Endocrinology (Baltimore, Md.)*, 25(10):1773–1785, 2011.

Delahaye, L., Rocchi, S., Van Obberghen, E. Potential involvement of FRS2 in insulin signaling. *Endocrinology*, 141(2):621–628, 2000.

Deng, C.X., Wynshaw-Boris, A., Shen, M.M., Daugherty, C., Ornitz, D.M., Leder, P. Murine FGFR-1 is required for early postimplantation growth and axial organization. *Genes & Development*, 8(24):3045–3057, 1994.

Dengl, S., Mayer, A., Sun, M., Cramer, P. Structure and in vivo requirement of the yeast Spt6 SH2 domain. *Journal of Molecular Biology*, 389(1):211–225, 2009.

Deribe, Y.L., Pawson, T., Dikic, I. Post-translational modifications in signal integration. *Nature Structural & Molecular Biology*, 17(6):666–672, 2010.

Desai, D.M., Sap, J., Schlessinger, J., Weiss, A. Ligand-mediated negative regulation of a chimeric transmembrane receptor tyrosine phosphatase. *Cell*, 73(3):541–554, 1993.

Dikic, I., Szymkiewicz, I., Soubeyran, P. Cbl signaling networks in the regulation of cell function. *Cellular and molecular life sciences: CMLS*, 60(9):1805–1827, 2003.

DiNitto, J.P., Lambright, D.G. Membrane and juxtamembrane targeting by PH and PTB domains. *Biochimica Et Biophysica Acta*, 1761(8):850–867, 2006.

Dombrosky-Ferlan, P.M., Corey, S.J. Yeast two-hybrid in vivo association of the Src kinase Lyn with the proto-oncogene product Cbl but not with the p85 subunit of PI 3-kinase. *Oncogene*, 14(17):2019–2024, 1997.

D'Oro, U., Ashwell, J.D. Cutting edge: the CD45 tyrosine phosphatase is an inhibitor of Lck activity in thymocytes. *Journal of Immunology (Baltimore, Md.: 1950)*, 162(4):1879–1883, 1999.

Downward, J., Parker, P., Waterfield, M.D. Autophosphorylation sites on the epidermal growth factor receptor. *Nature*, 311(5985):483–485, 1984.

Drappa, J., Kamen, L.A., Chan, E., Georgiev, M., Ashany, D., Marti, F., King, P.D. Impaired T cell death and lupus-like autoimmunity in T cell-specific adapter protein-deficient mice. *The Journal of Experimental Medicine*, 198(5):809–821, 2003.

Drouin, G., Moniz de Sá, M., Zuker, M. The Giardia lamblia actin gene and the phylogeny of eukaryotes. *Journal of Molecular Evolution*, 41(6):841–849, 1995.

Duffy, S.L., Coulthard, M.G., Spanevello, M.D., Herath, N.I., Yeadon, T.M., McCarron, J.K., Carter, J.C., Tonks, I.D., Kay, G.F., Phillips, G.E., Boyd, A.W. Generation and characterization of EphA1 receptor tyrosine kinase reporter knockout mice. *Genesis (New York, N.Y.: 2000)*, 46(10):553–561, 2008.

Dumont, D.J., Gradwohl, G., Fong, G.H., Puri, M.C., Gertsenstein, M., Auerbach, A., Breitman, M.L. Dominant-negative and targeted null mutations in the endothelial receptor tyrosine kinase, tek, reveal a critical role in vasculogenesis of the embryo. *Genes & Development*, 8(16):1897–1909, 1994.

Dumont, D.J., Jussila, L., Taipale, J., Lymboussaki, A., Mustonen, T., Pajusola, K., Breitman, M., Alitalo, K. Cardiovascular failure in mouse embryos deficient in VEGF receptor-3. *Science (New York, N.Y.)*, 282(5390):946–949, 1998.

Duncan, J.L., LaVail, M.M., Yasumura, D., Matthes, M.T., Yang, H., Trautmann, N., Chappelow, A.V., Feng, W., Earp, H.S., Matsushima, G.K., Vollrath, D. An RCS-like retinal dystrophy phenotype in mer knockout mice. *Investigative Ophthalmology & Visual Science*, 44(2):826–838, 2003.

Dyson, N. The regulation of E2F by pRB-family proteins. *Genes & Development*, 12(15):2245–2262, 1998.

Eder, J.P., Vande Woude, G.F., Boerner, S.A., LoRusso, P.M. Novel therapeutic inhibitors of the c-Met signaling pathway in cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 15(7):2207–2214, 2009.

Eigen, M., Schuster, P. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Die Naturwissenschaften*, 64(11):541–565, 1977.

Elly, C., Witte, S., Zhang, Z., Rosnet, O., Lipkowitz, S., Altman, A., Liu, Y.C. Tyrosine phosphorylation and complex formation of Cbl-b upon T cell receptor stimulation. *Oncogene*, 18(5):1147–1156, 1999.

Errede, B., Gartner, A., Zhou, Z., Nasmyth, K., Ammerer, G. MAP kinase-related FUS3 from S. cerevisiae is activated by STE7 in vitro. *Nature*, 362(6417):261–264, 1993.

Farooq, A., Zeng, L., Yan, K.S., Ravichandran, K.S., Zhou, M.M. Coupling of folding and binding in the PTB domain of the signaling protein Shc. *Structure (London, England: 1993)*, 11(8):905–913, 2003.

Fazili, Z., Sun, W., Mittelstaedt, S., Cohen, C., Xu, X.X. Disabled-2 inactivation is an early step in ovarian tumorigenicity. *Oncogene*, 18(20):3104–3113, 1999.

Felder, S., Zhou, M., Hu, P., Ureña, J., Ullrich, A., Chaudhuri, M., White, M., Shoelson, S.E., Schlessinger, J. SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange. *Molecular and Cellular Biology*, 13(3):1449–1455, 1993.

Feldheim, D.A., Nakamoto, M., Osterfield, M., Gale, N.W., DeChiara, T.M., Rohatgi, R., Yancopoulos, G.D., Flanagan, J.G. Loss-of-function analysis of EphA receptors in retinotectal mapping. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 24(10):2542–2550, 2004.

Fenyö, D. Identifying the proteome: software tools. *Current Opinion in Biotechnology*, 11(4):391–395, 2000.

Fields, S., Song, O. A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246, 1989.

Florio, M., Wilson, L.K., Trager, J.B., Thorner, J., Martin, G.S. Aberrant protein phosphorylation at tyrosine is responsible for the growth-inhibitory action of pp60v-src expressed in the yeast Saccharomyces cerevisiae. *Molecular Biology of the Cell*, 5(3):283–296, 1994.

Fong, G.H., Rossant, J., Gertsenstein, M., Breitman, M.L. Role of the Flt-1 receptor tyrosine kinase in regulating the assembly of vascular endothelium. *Nature*, 376(6535):66–70, 1995.

Fournier, E., Isakoff, S.J., Ko, K., Cardinale, C.J., Inghirami, G.G., Li, Z., Curotto de Lafaille, M.A., Skolnik, E.Y. The B cell SH2/PH domain-containing adaptor Bam32/DAPP1 is required for T cell-independent II antigen responses. *Current biology: CB*, 13(21):1858–1866, 2003.

Frame, M.C. Src in cancer: deregulation and consequences for cell behaviour. *Biochimica Et Biophysica Acta*, 1602(2):114–130, 2002.

Frey, B.J., Dueck, D. Clustering by passing messages between data points. *Science (New York, N.Y.)*, 315(5814):972–976, 2007.

Fujita, Y., Krause, G., Scheffner, M., Zechner, D., Leddy, H.E.M., Behrens, J., Sommer, T., Birchmeier, W. Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. *Nature Cell Biology*, 4(3):222–231, 2002.

Fukazawa, T., Miyake, S., Band, V., Band, H. Tyrosine phosphorylation of Cbl upon epidermal growth factor (EGF) stimulation and its association with EGF receptor and downstream signaling proteins. *The Journal of Biological Chemistry*, 271(24):14554–14559, 1996.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., Stratton, M.R. A census of human cancer genes. *Nature Reviews. Cancer*, 4(3):177–183, 2004.

Gadue, P., Morton, N., Stein, P.L. The Src family tyrosine kinase Fyn regulates natural killer T cell development. *The Journal of Experimental Medicine*, 190(8):1189–1196, 1999.

Gassmann, M., Casagranda, F., Orioli, D., Simon, H., Lai, C., Klein, R., Lemke, G. Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. *Nature*, 378(6555):390–394, 1995.

Gavin, A.C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G.,

Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., Superti-Furga, G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415(6868):141–147, 2002.

Gerety, S.S., Wang, H.U., Chen, Z.F., Anderson, D.J. Symmetrical mutant phenotypes of the receptor EphB4 and its specific transmembrane ligand ephrin-B2 in cardiovascular development. *Molecular Cell*, 4(3):403–414, 1999.

Gerwien, J., Nielsen, M., Labuda, T., Nissen, M.H., Svejgaard, A., Geisler, C., Röpke, C., Odum, N. Cutting edge: TCR stimulation by antibody and bacterial superantigen induces Stat3 activation in human T cells. *Journal of Immunology (Baltimore, Md.: 1950)*, 163(4):1742–1745, 1999.

Gery, S., Gueller, S., Nowak, V., Sohn, J., Hofmann, W.K., Koeffler, H.P. Expression of the adaptor protein Lnk in leukemia cells. *Experimental Hematology*, 37(5):585–592.e2, 2009.

Gietz, R.D., Woods, R.A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods in Enzymology*, 350:87–96, 2002.

Gil, D., Schamel, W.W.A., Montoya, M., Sánchez-Madrid, F., Alarcón, B. Recruitment of Nck by CD3 epsilon reveals a ligand-induced conformational change essential for T cell receptor signaling and synapse formation. *Cell*, 109(7):901–912, 2002.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., Vijayadamodar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioime, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carrolla, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C.A., Finley, R.L., White, K.P., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R.A., McKenna, M.P., Chant, J., Rothberg, J.M. A protein interaction map of Drosophila melanogaster. *Science (New York, N.Y.)*, 302(5651):1727–1736, 2003.

Gnad, F., de Godoy, L.M.F., Cox, J., Neuhauser, N., Ren, S., Olsen, J.V., Mann, M. High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. *Proteomics*, 9(20):4642–4652, 2009.

Goehler, H., Lalowski, M., Stelzl, U., Waelter, S., Stroedicke, M., Worm, U., Droege, A., Lindenberg, K.S., Knoblich, M., Haenig, C., Herbst, M., Suopanki, J., Scherzinger, E., Abraham, C., Bauer, B., Hasenbank, R., Fritzsche, A., Ludewig, A.H., Büssow, K., Buessow, K., Coleman, S.H., Gutekunst, C.A., Landwehrmeyer, B.G., Lehrach, H., Wanker, E.E. A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Molecular Cell*, 15(6):853–865, 2004.

Goh, E.L.K., Zhu, T., Leong, W.Y., Lobie, P.E. c-Cbl is a negative regulator of GH-stimulated STAT5-mediated transcription. *Endocrinology*, 143(9):3590–3603, 2002.

Goitsuka, R., Kanazashi, H., Sasanuma, H., Fujimura, Y., Hidaka, Y., Tatsuno, A., Ra, C., Hayashi, K., Kitamura, D. A BASH/SLP-76-related adaptor protein MIST/Clnk involved in IgE receptor-mediated mast cell degranulation. *International Immunology*, 12(4):573–580, 2000.

Gong, Q., Cheng, A.M., Akk, A.M., Alberola-Ila, J., Gong, G., Pawson, T., Chan, A.C. Disruption of T cell signaling networks and development by Grb2 haploid insufficiency. *Nature Immunology*, 2(1):29–36, 2001.

Graham, F.L., Smiley, J., Russell, W.C., Nairn, R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *The Journal of General Virology*, 36(1):59–74, 1977.

Granum, S., Andersen, T.C.B., Sørlie, M., Jørgensen, M., Koll, L., Berge, T., Lea, T., Fleckenstein, B., Spurkland, A., Sundvold-Gjerstad, V. Modulation of Lck function through multisite docking to T cell-specific adapter protein. *The Journal of Biological Chemistry*, 283(32):21909–21919, 2008.

Graus-Porta, D., Beerli, R.R., Daly, J.M., Hynes, N.E. ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *The EMBO journal*, 16(7):1647–1655, 1997.

Greaves, M., Maley, C.C. Clonal evolution in cancer. *Nature*, 481(7381):306–313, 2012.

de Groot, R.P., Raaijmakers, J.A., Lammers, J.W., Jove, R., Koenderman, L. STAT5 activation by BCR-Abl contributes to transformation of K562 leukemia cells. *Blood*, 94(3):1108–1112, 1999.

Grosberg, R.K., Strathmann, R.R. The evolution of multicellularity: a minor major transition? *Annual Review of Ecology, Evolution, and Systematics*, pp. 621–654, 2007.

Grossmann, A., Benlasfer, N., Birth, P., Hegele, A., Wachsmuth, F., Apelt, L., Stelzl, U. Phosphotyrosine-dependent protein-protein interaction network. *Molecular Systems Biology*, 11:794, 2015.

Grossmann, A.H., Kolibaba, K.S., Willis, S.G., Corbin, A.S., Langdon, W.S., Deininger, M.W.N., Druker, B.J. Catalytic domains of tyrosine kinases determine the phosphorylation sites within c-Cbl. *FEBS letters*, 577(3):555–562, 2004.

Grøvdal, L.M., Stang, E., Sorkin, A., Madshus, I.H. Direct interaction of Cbl with pTyr 1045 of the EGF receptor (EGFR) is required to sort the EGFR to lysosomes for degradation. *Experimental Cell Research*, 300(2):388–395, 2004.

Gschwind, A., Fischer, O.M., Ullrich, A. The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nature Reviews. Cancer*, 4(5):361–370, 2004.

Gueller, S., Hehn, S., Nowak, V., Gery, S., Serve, H., Brandts, C.H., Koeffler, H.P. Adaptor protein Lnk binds to PDGF receptor and inhibits PDGF-dependent signaling. *Experimental Hematology*, 39(5):591–600, 2011.

Guy, P.M., Platko, J.V., Cantley, L.C., Cerione, R.A., Carraway, K.L. Insect cell-expressed p180erbB3 possesses an impaired tyrosine kinase activity. *Proceedings of the National Academy of Sciences of the United States of America*, 91(17):8132–8136, 1994.

Hackel, P.O., Zwick, E., Prenzel, N., Ullrich, A. Epidermal growth factor receptors: critical mediators of multiple receptor pathways. *Current Opinion in Cell Biology*, 11(2):184–189, 1999.

Halford, M.M., Armes, J., Buchert, M., Meskenaite, V., Grail, D., Hibbs, M.L., Wilks, A.F., Farlie, P.G., Newgreen, D.F., Hovens, C.M., Stacker, S.A. Ryk-deficient mice exhibit craniofacial defects associated with perturbed Eph receptor crosstalk. *Nature Genetics*, 25(4):414–418, 2000.

Han, J.H., Batey, S., Nickson, A.A., Teichmann, S.A., Clarke, J. The folding and evolution of multidomain proteins. *Nature Reviews. Molecular Cell Biology*, 8(4):319–330, 2007.

Hanahan, D., Weinberg, R.A. The hallmarks of cancer. *Cell*, 100(1):57–70, 2000.

Hanke, S., Mann, M. The phosphotyrosine interactome of the insulin receptor family and its substrates IRS-1 and IRS-2. *Molecular & cellular proteomics: MCP*, 8(3):519–534, 2009.

Hanks, S.K., Ryzhova, L., Shin, N.Y., Brábek, J. Focal adhesion kinase signaling activities and their implications in the control of cell survival and motility. *Frontiers in Bioscience: A Journal and Virtual Library*, 8:d982–996, 2003.

Harr, M.W., Caimi, P.F., McColl, K.S., Zhong, F., Patel, S.N., Barr, P.M., Distelhorst, C.W. Inhibition of Lck enhances glucocorticoid sensitivity and apoptosis in lymphoid cell lines and in chronic lymphocytic leukemia. *Cell Death and Differentiation*, 17(9):1381–1391, 2010.

Harris, C.C. p53 tumor suppressor gene: from the basic research laboratory to the clinic–an abridged historical perspective. *Carcinogenesis*, 17(6):1187–1198, 1996.

Harrison, S.C. Peptide-surface association: the case of PDZ and PTB domains. *Cell*, 86(3):341–343, 1996.

Harrison, S.C. Variation on an Src-like theme. *Cell*, 112(6):737–740, 2003.

Hause, R.J., Leung, K.K., Barkinge, J.L., Ciaccio, M.F., Chuu, C.P., Jones, R.B. Comprehensive binary interaction mapping of SH2 domains via fluorescence polarization reveals novel functional diversification of ErbB receptors. *PloS One*, 7(9):e44471, 2012.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., Stelzl, U. Dynamic protein-protein interaction wiring of the human spliceosome. *Molecular Cell*, 45(4):567–580, 2012.

Helin, K., Harlow, E., Fattaey, A. Inhibition of E2F-1 transactivation by direct binding of the retinoblastoma protein. *Molecular and Cellular Biology*, 13(10):6501–6508, 1993.

Helmbacher, F., Schneider-Maunoury, S., Topilko, P., Tiret, L., Charnay, P. Targeting of the EphA4 tyrosine kinase receptor affects dorsal/ventral pathfinding of limb motor axons. *Development (Cambridge, England)*, 127(15):3313–3324, 2000.

Herrin, B.R., Groeger, A.L., Justement, L.B. The adaptor protein HSH2 attenuates apoptosis in response to ligation of the B cell antigen receptor complex on the B lymphoma cell line, WEHI-231. *The Journal of Biological Chemistry*, 280(5):3507–3515, 2005.

Hildeman, D.A., Zhu, Y., Mitchell, T.C., Bouillet, P., Strasser, A., Kappler, J., Marrack, P. Activated T cell death in vivo mediated by proapoptotic bcl-2 family member bim. *Immunity*, 16(6):759–767, 2002a.

Hildeman, D.A., Zhu, Y., Mitchell, T.C., Kappler, J., Marrack, P. Molecular mechanisms of activated T cell death in vivo. *Current Opinion in Immunology*, 14(3):354–359, 2002b.

Himpe, E., Kooijman, R. Insulin-like growth factor-I receptor signal transduction and the Janus Kinase/Signal Transducer and Activator of Transcription (JAK-STAT) pathway. *BioFactors (Oxford, England)*, 35(1):76–81, 2009.

Hino, N., Oyama, M., Sato, A., Mukai, T., Iraha, F., Hayashi, A., Kozuka-Hata, H., Yamamoto, T., Yokoyama, S., Sakamoto, K. Genetic incorporation of a

photo-crosslinkable amino acid reveals novel protein complexes with GRB2 in mammalian cells. *Journal of Molecular Biology*, 406(2):343–353, 2011.

Hirano, T., Ishihara, K., Hibi, M. Roles of STAT3 in mediating the cell growth, differentiation and survival signals relayed through the IL-6 family of cytokine receptors. *Oncogene*, 19(21):2548–2556, 2000.

Hoenigsberg, H.F., Tíjaro, M.H., Sanabria, C. From unicellularity to multicellularity - molecular speculations about early animal evolution. *Genetics and molecular research: GMR*, 7(1):50–59, 2008.

Hof, P., Pluskey, S., Dhe-Paganon, S., Eck, M.J., Shoelson, S.E. Crystal structure of the tyrosine phosphatase SHP-2. *Cell*, 92(4):441–450, 1998.

Holland, S.J., Liao, X.C., Mendenhall, M.K., Zhou, X., Pardo, J., Chu, P., Spencer, C., Fu, A., Sheng, N., Yu, P., Pali, E., Nagin, A., Shen, M., Yu, S., Chan, E., Wu, X., Li, C., Woisetschlager, M., Aversa, G., Kolbinger, F., Bennett, M.K., Molineaux, S., Luo, Y., Payan, D.G., Mancebo, H.S., Wu, J. Functional cloning of Src-like adapter protein-2 (SLAP-2), a novel inhibitor of antigen receptor signaling. *The Journal of Experimental Medicine*, 194(9):1263–1276, 2001.

Holmqvist, K., Cross, M.J., Rolny, C., Hägerkvist, R., Rahimi, N., Matsumoto, T., Claesson-Welsh, L., Welsh, M. The adaptor protein shb binds to tyrosine 1175 in vascular endothelial growth factor (VEGF) receptor-2 and regulates VEGF-dependent cellular migration. *The Journal of Biological Chemistry*, 279(21):22267–22275, 2004.

Hom, G., Graham, R.R., Modrek, B., Taylor, K.E., Ortmann, W., Garnier, S., Lee, A.T., Chung, S.A., Ferreira, R.C., Pant, P.V.K., Ballinger, D.G., Kosoy, R., Demirci, F.Y., Kamboh, M.I., Kao, A.H., Tian, C., Gunnarsson, I., Bengtsson, A.A., Rantapää-Dahlqvist, S., Petri, M., Manzi, S., Seldin, M.F., Rönnblom, L., Syvänen, A.C., Criswell, L.A., Gregersen, P.K., Behrens, T.W. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *The New England Journal of Medicine*, 358(9):900–909, 2008.

Hong, W.K., Hait, W. *Holland Frei Cancer Medicine*, volume 8. PMPH-USA, 2010.

Horn, S.C., Lalowski, M., Goehler, H., Dröge, A., Wanker, E.E., Stelzl, U. Huntingtin interacts with the receptor sorting family protein GASP2. *Journal of Neural Transmission (Vienna, Austria: 1996)*, 113(8):1081–1090, 2006.

Hou, S., Pauls, S.D., Liu, P., Marshall, A.J. The PH domain adaptor protein Bam32/DAPP1 functions in mast cells to restrain FcϵRI-induced calcium flux and granule release. *Molecular Immunology*, 48(1-3):89–97, 2010.

Howard, P.L., Chia, M.C., Del Rizzo, S., Liu, F.F., Pawson, T. Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 100(20):11267–11272, 2003.

Howlader, N., Noone, A., Krapcho, M., Neyman, N., Aminou, R., Waldron, W., Altekruse, S., Kosary, C., Ruhl, J., Tatalovich, Z., Cho, H., Mariotto, A., Eisner, M., Lewis, D., Chen, H., Feuer, E., Cronin, K. SEER cancer statistics review, 1975–2009 (vintage 2009 populations). *Bethesda, MD: National Cancer Institute*, 2012.

Howlett, C.J., Bisson, S.A., Resek, M.E., Tigley, A.W., Robbins, S.M. The proto-oncogene p120(Cbl) is a downstream substrate of the Hck protein-tyrosine kinase. *Biochemical and Biophysical Research Communications*, 257(1):129–138, 1999.

Huang, C.Y., Tan, T.H. DUSPs, to MAP kinases and beyond. *Cell & Bioscience*, 2(1):24, 2012.

Huang, H., Li, L., Wu, C., Schibli, D., Colwill, K., Ma, S., Li, C., Roy, P., Ho, K., Songyang, Z., Pawson, T., Gao, Y., Li, S.S.C. Defining the specificity space of the human SRC homology 2 domain. *Molecular & cellular proteomics: MCP*, 7(4):768–784, 2008.

Hubbard, S.R., Miller, W.T. Receptor tyrosine kinases: mechanisms of activation and signaling. *Current Opinion in Cell Biology*, 19(2):117–123, 2007.

Hubbard, S.R., Till, J.H. Protein tyrosine kinase structure and function. *Annual Review of Biochemistry*, 69:373–398, 2000.

Hunter, S., Burton, E.A., Wu, S.C., Anderson, S.M. Fyn associates with Cbl and phosphorylates tyrosine 731 in Cbl, a binding site for phosphatidylinositol 3-kinase. *The Journal of Biological Chemistry*, 274(4):2097–2106, 1999.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., de Castro, E., Coggill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R.D., Fraser, M., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., McMenamin, C., Mi, H., Mutowo-Muellenet, P., Mulder, N., Natale, D., Orengo, C., Pesseat, S., Punta, M., Quinn, A.F., Rivoire, C., Sangrador-Vegas, A., Selengut, J.D., Sigrist, C.J.A., Scheremetjew, M., Tate, J., Thimmajanarthanan, M., Thomas, P.D., Wu, C.H., Yeats, C., Yong, S.Y. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research*, 40(Database issue):D306–312, 2012.

Hunter, T. Tyrosine phosphorylation: thirty years and counting. *Current Opinion in Cell Biology*, 21(2):140–146, 2009.

Hunter, T., Cooper, J.A. Protein-tyrosine kinases. *Annual Review of Biochemistry*, 54:897–930, 1985.

Hunter, T., Sefton, B.M. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proceedings of the National Academy of Sciences of the United States of America*, 77(3):1311–1315, 1980.

Ilaria, R.L., Van Etten, R.A. P210 and P190(BCR/ABL) induce the tyrosine phosphorylation and DNA binding activity of multiple specific STAT family members. *The Journal of Biological Chemistry*, 271(49):31704–31710, 1996.

Ingley, E., Schneider, J.R., Payne, C.J., McCarthy, D.J., Harder, K.W., Hibbs, M.L., Klinken, S.P. Csk-binding protein mediates sequential enzymatic down-regulation and degradation of Lyn in erythropoietin-stimulated cells. *The Journal of Biological Chemistry*, 281(42):31920–31929, 2006.

Iseki, M., Kubo, C., Kwon, S.M., Yamaguchi, A., Kataoka, Y., Yoshida, N., Takatsu, K., Takaki, S. Increased numbers of B-1 cells and enhanced responses against TI-2 antigen in mice lacking APS, an adaptor molecule containing PH and SH2 domains. *Molecular and Cellular Biology*, 24(6):2243–2250, 2004.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences of the United States of America*, 98(8):4569–4574, 2001.

Jeon, M.S., Atfield, A., Venuprasad, K., Krawczyk, C., Sarao, R., Elly, C., Yang, C., Arya, S., Bachmaier, K., Su, L., Bouchard, D., Jones, R., Gronski, M., Ohashi, P., Wada, T., Bloom, D., Fathman, C.G., Liu, Y.C., Penninger, J.M. Essential role of the E3 ubiquitin ligase Cbl-b in T cell anergy induction. *Immunity*, 21(2):167–177, 2004.

Jones, R.B., Gordus, A., Krall, J.A., MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature*, 439(7073):168–174, 2006.

Jonsson, P.F., Bates, P.A. Global topological features of cancer proteins in the human interactome. *Bioinformatics (Oxford, England)*, 22(18):2291–2297, 2006.

Joshi, R.L., Lamothe, B., Cordonnier, N., Mesbah, K., Monthioux, E., Jami, J., Bucchini, D. Targeted disruption of the insulin receptor gene in the mouse results in neonatal lethality. *The EMBO journal*, 15(7):1542–1547, 1996.

Joyce, G.F. The antiquity of RNA-based evolution. *Nature*, 418(6894):214–221, 2002.

Jumaa, H., Bossaller, L., Portugal, K., Storch, B., Lotz, M., Flemming, A., Schrappe, M., Postila, V., Riikonen, P., Pelkonen, J., Niemeyer, C.M., Reth, M. Deficiency of the adaptor SLP-65 in pre-B-cell acute lymphoblastic leukaemia. *Nature*, 423(6938):452–456, 2003.

Kaake, R.M., Wang, X., Huang, L. Profiling of protein interaction networks of protein complexes using affinity purification and quantitative mass spectrometry. *Molecular & cellular proteomics: MCP*, 9(8):1650–1665, 2010.

Kacinski, B.M. CSF-1 and its receptor in ovarian, endometrial and breast cancer. *Annals of Medicine*, 27(1):79–85, 1995.

Kadota, S., Fantus, I.G., Deragon, G., Guyda, H.J., Posner, B.I. Stimulation of insulin-like growth factor II receptor binding and insulin receptor kinase activity in rat adipocytes. Effects of vanadate and H2O2. *The Journal of Biological Chemistry*, 262(17):8252–8256, 1987.

Kaiser, D. Building a multicellular organism. *Annual Review of Genetics*, 35:103–123, 2001.

Kamburov, A. *More complete and more accurate interactomes for elucidating the mechanisms of complex diseases*. Berlin, 2012.

Kamburov, A., Cavill, R., Ebbels, T.M.D., Herwig, R., Keun, H.C. Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. *Bioinformatics (Oxford, England)*, 27(20):2917–2918, 2011.

Kamburov, A., Grossmann, A., Herwig, R., Stelzl, U. Cluster-based assessment of protein-protein interaction confidence. *BMC bioinformatics*, 13:262, 2012.

Kamburov, A., Wierling, C., Lehrach, H., Herwig, R. ConsensusPathDB–a database for integrating human functional interaction networks. *Nucleic Acids Research*, 37(Database issue):D623–628, 2009.

Kamps, M.P., Taylor, S.S., Sefton, B.M. Direct evidence that oncogenic tyrosine kinases and cyclic AMP-dependent protein kinase have homologous ATP-binding sites. *Nature*, 310(5978):589–592, 1984.

Kandasamy, K., Mohan, S.S., Raju, R., Keerthikumar, S., Kumar, G.S.S., Venugopal, A.K., Telikicherla, D., Navarro, J.D., Mathivanan, S., Pecquet, C., Gollapudi, S.K., Tattikota, S.G., Mohan, S., Padhukasahasram, H., Subbannayya, Y., Goel, R., Jacob, H.K.C., Zhong, J., Sekhar, R., Nanjappa, V., Balakrishnan, L., Subbaiah, R., Ramachandra, Y.L., Rahiman, B.A., Prasad,

T.S.K., Lin, J.X., Houtman, J.C.D., Desiderio, S., Renauld, J.C., Constantinescu, S.N., Ohara, O., Hirano, T., Kubo, M., Singh, S., Khatri, P., Draghici, S., Bader, G.D., Sander, C., Leonard, W.J., Pandey, A. NetPath: a public resource of curated signal transduction pathways. *Genome Biology*, 11(1):R3, 2010.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, 40(Database issue):D109–114, 2012.

Kaneko, T., Huang, H., Cao, X., Li, X., Li, C., Voss, C., Sidhu, S.S., Li, S.S.C. Superbinder SH2 domains act as antagonists of cell signaling. *Science Signaling*, 5(243):ra68, 2012a.

Kaneko, T., Joshi, R., Feller, S.M., Li, S.S. Phosphotyrosine recognition domains: the typical, the atypical and the versatile. *Cell communication and signaling: CCS*, 10(1):32, 2012b.

Kashtan, N., Alon, U. Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39):13773–13778, 2005.

Katso, R.M., Russell, R.B., Ganesan, T.S. Functional analysis of H-Ryk, an atypical member of the receptor tyrosine kinase family. *Molecular and Cellular Biology*, 19(9):6427–6440, 1999.

Kawata, T., Shevchenko, A., Fukuzawa, M., Jermyn, K.A., Totty, N.F., Zhukovskaya, N.V., Sterling, A.E., Mann, M., Williams, J.G. SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in dictyostelium. *Cell*, 89(6):909–916, 1997.

Keegan, K., Cooper, J.A. Use of the two hybrid system to detect the association of the protein-tyrosine-phosphatase, SHPTP2, with another SH2-containing protein, Grb7. *Oncogene*, 12(7):1537–1544, 1996.

Kelder, T., van Iersel, M.P., Hanspers, K., Kutmon, M., Conklin, B.R., Evelo, C.T., Pico, A.R. WikiPathways: building research communities on biological pathways. *Nucleic Acids Research*, 40(Database issue):D1301–1307, 2012.

Kennelly, P.J. Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. *FEMS microbiology letters*, 206(1):1–8, 2002.

Kennelly, P.J., Potts, M. Fancy meeting you here! A fresh look at "prokaryotic"protein phosphorylation. *Journal of Bacteriology*, 178(16):4759–4764, 1996.

Kerppola, T.K. Visualization of molecular interactions by fluorescence complementation. *Nature Reviews. Molecular Cell Biology*, 7(6):449–456, 2006.

Keskin, O., Tsai, C.J., Wolfson, H., Nussinov, R. A new, structurally nonredundant, diverse data set of protein-protein interfaces and its implications. *Protein Science: A Publication of the Protein Society*, 13(4):1043–1055, 2004.

Keyse, S.M. Protein phosphatases and the regulation of mitogen-activated protein kinase signalling. *Current Opinion in Cell Biology*, 12(2):186–192, 2000.

Kile, B.T., Nicola, N.A., Alexander, W.S. Negative regulators of cytokine signaling. *International Journal of Hematology*, 73(3):292–298, 2001.

King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K.J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J.B., Morris, A., Nichols, S., Richter, D.J., Salamov, A., Sequencing, J.G.I., Bork, P., Lim, W.A., Manning, G., Miller, W.T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I.V., Rokhsar, D. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature*, 451(7180):783–788, 2008.

Kinzler, K.W., Vogelstein, B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature*, 386(6627):761, 763, 1997.

Kisseleva, T., Bhattacharya, S., Braunstein, J., Schindler, C.W. Signaling through the JAK/STAT pathway, recent advances and future challenges. *Gene*, 285(1-2):1–24, 2002.

Kitamura, T., Kido, Y., Nef, S., Merenmies, J., Parada, L.F., Accili, D. Preserved pancreatic beta-cell development and function in mice lacking the insulin

receptor-related receptor. *Molecular and Cellular Biology*, 21(16):5624–5630, 2001.

Kitano, H. Biological robustness. *Nature Reviews. Genetics*, 5(11):826–837, 2004.

Kito, K., Kawaguchi, N., Okada, S., Ito, T. Discrimination between stable and dynamic components of protein complexes by means of quantitative proteomics. *Proteomics*, 8(12):2366–2370, 2008.

Klapper, L.N., Glathe, S., Vaisman, N., Hynes, N.E., Andrews, G.C., Sela, M., Yarden, Y. The ErbB-2/HER2 oncoprotein of human carcinomas may function solely as a shared coreceptor for multiple stroma-derived growth factors. *Proceedings of the National Academy of Sciences of the United States of America*, 96(9):4995–5000, 1999.

Klein, R., Silos-Santiago, I., Smeyne, R.J., Lira, S.A., Brambilla, R., Bryant, S., Zhang, L., Snider, W.D., Barbacid, M. Disruption of the neurotrophin-3 receptor gene trkC eliminates la muscle afferents and results in abnormal movements. *Nature*, 368(6468):249–251, 1994.

Klein, R., Smeyne, R.J., Wurst, W., Long, L.K., Auerbach, B.A., Joyner, A.L., Barbacid, M. Targeted disruption of the trkB neurotrophin receptor gene results in nervous system lesions and neonatal death. *Cell*, 75(1):113–122, 1993.

Kolltveit, K.M., Granum, S., Aasheim, H.C., Forsbring, M., Sundvold-Gjerstad, V., Dai, K.Z., Molberg, O., Schjetne, K.W., Bogen, B., Shapiro, V.S., Johansen, F.E., Schenck, K., Spurkland, A. Expression of SH2D2A in T-cells is regulated both at the transcriptional and translational level. *Molecular Immunology*, 45(8):2380–2390, 2008.

Kolltveit, K.M., Schreurs, O., Ostrem, J., Søland, T.M., Khuu, C., Berge, T., Messelt, E., Hayashi, K., Granum, S., Spurkland, A., Schenck, K. Expression of the T-cell-specific adapter protein in oral epithelium. *European Journal of Oral Sciences*, 118(2):159–167, 2010.

Koonin, E.V., Aravind, L., Kondrashov, A.S. The impact of comparative genomics on our understanding of evolution. *Cell*, 101(6):573–576, 2000.

Kotani, K., Wilden, P., Pillay, T.S. SH2-Balpha is an insulin-receptor adapter protein and substrate that interacts with the activation loop of the insulin-receptor kinase. *The Biochemical Journal*, 335 (Pt 1):103–109, 1998.

Kouhara, H., Hadari, Y.R., Spivak-Kroizman, T., Schilling, J., Bar-Sagi, D., Lax, I., Schlessinger, J. A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway. *Cell*, 89(5):693–702, 1997.

von Kriegsheim, A., Baiocchi, D., Birtwistle, M., Sumpton, D., Bienvenut, W., Morrice, N., Yamada, K., Lamond, A., Kalna, G., Orton, R., Gilbert, D., Kolch, W. Cell fate decisions are specified by the dynamic ERK interactome. *Nature Cell Biology*, 11(12):1458–1464, 2009.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., Punna, T., Peregrín-Alvarez, J.M., Shales, M., Zhang, X., Davey, M., Robinson, M.D., Paccanaro, A., Bray, J.E., Sheung, A., Beattie, B., Richards, D.P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M.M., Vlasblom, J., Wu, S., Orsi, C., Collins, S.R., Chandran, S., Haw, R., Rilstone, J.J., Gandi, K., Thompson, N.J., Musso, G., St Onge, P., Ghanny, S., Lam, M.H.Y., Butland, G., Altaf-Ul, A.M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J.S., Ingles, C.J., Hughes, T.R., Parkinson, J., Gerstein, M., Wodak, S.J., Emili, A., Greenblatt, J.F. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature*, 440(7084):637–643, 2006.

Kuriyan, J., Cowburn, D. Modular peptide recognition domains in eukaryotic signaling. *Annual Review of Biophysics and Biomolecular Structure*, 26:259–288, 1997.

Labrador, J.P., Azcoitia, V., Tuckermann, J., Lin, C., Olaso, E., Mañes, S., Brückner, K., Goergen, J.L., Lemke, G., Yancopoulos, G., Angel, P., Martínez, C., Klein, R. The collagen receptor DDR2 regulates proliferation and its elimination leads to dwarfism. *EMBO reports*, 2(5):446–452, 2001.

Laird, R.M., Laky, K., Hayes, S.M. Unexpected role for the B cell-specific Src family kinase B lymphoid kinase in the development of IL-17-producing $\gamma\delta$

T cells. *Journal of Immunology (Baltimore, Md.: 1950)*, 185(11):6518–6527, 2010.

Laity, J.H., Lee, B.M., Wright, P.E. Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology*, 11(1):39–46, 2001.

Lamalice, L., Houle, F., Jourdan, G., Huot, J. Phosphorylation of tyrosine 1214 on VEGFR2 is required for VEGF-induced activation of Cdc42 upstream of SAPK2/p38. *Oncogene*, 23(2):434–445, 2004.

Lamesch, P., Li, N., Milstein, S., Fan, C., Hao, T., Szabo, G., Hu, Z., Venkatesan, K., Bethel, G., Martin, P., Rogers, J., Lawlor, S., McLaren, S., Dricot, A., Borick, H., Cusick, M.E., Vandenhaute, J., Dunham, I., Hill, D.E., Vidal, M. hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics*, 89(3):307–315, 2007.

Lange, V., Picotti, P., Domon, B., Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology*, 4:222, 2008.

Lapinski, P.E., Oliver, J.A., Bodie, J.N., Marti, F., King, P.D. The T-cell-specific adapter protein family: TSAd, ALX, and SH2D4A/SH2D4B. *Immunological Reviews*, 232(1):240–254, 2009.

Lee, J., Andreeva, A., Sipe, C.W., Liu, L., Cheng, A., Lu, X. PTK7 regulates myosin II activity to orient planar polarity in the mammalian auditory epithelium. *Current biology: CB*, 22(11):956–966, 2012.

Lee, J., Pilch, P.F. The insulin receptor: structure, function, and signaling. *The American Journal of Physiology*, 266(2 Pt 1):C319–334, 1994.

Lee, J.R., Hahn, H.S., Kim, Y.H., Nguyen, H.H., Yang, J.M., Kang, J.S., Hahn, M.J. Adaptor protein containing PH domain, PTB domain and leucine zipper (APPL1) regulates the protein level of EGFR by modulating its trafficking. *Biochemical and Biophysical Research Communications*, 415(1):206–211, 2011.

Lee, K.F., Simon, H., Chen, H., Bates, B., Hung, M.C., Hauser, C. Requirement for neuregulin receptor erbB2 in neural and cardiac development. *Nature*, 378(6555):394–398, 1995.

Lee, M.R., Chung, C.S., Liou, M.L., Wu, M., Li, W.F., Hsueh, Y.P., Lai, M.Z. Isolation and characterization of nuclear proteins that bind to T cell receptor V beta decamer motif. *Journal of Immunology (Baltimore, Md.: 1950)*, 148(6):1906–1912, 1992.

Lehner, B., Fraser, A.G. A first-draft human protein-interaction map. *Genome Biology*, 5(9):R63, 2004.

Lemmon, M.A., Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell*, 141(7):1117–1134, 2010.

Levinson, N.M., Seeliger, M.A., Cole, P.A., Kuriyan, J. Structural basis for the recognition of c-Src by its inactivator Csk. *Cell*, 134(1):124–134, 2008.

Levy, D.E., Lee, C. What does Stat3 do? *The Journal of Clinical Investigation*, 109(9):1143–1148, 2002.

Li, B.Q., Huang, T., Liu, L., Cai, Y.D., Chou, K.C. Identification of colorectal cancer related genes with mRMR and shortest path in protein-protein interaction network. *PloS One*, 7(4):e33393, 2012.

Li, J., Mao, X., Dong, L.Q., Liu, F., Tong, L. Crystal structures of the BAR-PH and PTB domains of human APPL1. *Structure (London, England: 1993)*, 15(5):525–533, 2007.

Li, L., Wu, C., Huang, H., Zhang, K., Gan, J., Li, S.S.C. Prediction of phosphotyrosine signaling networks using a scoring matrix-assisted ligand identification approach. *Nucleic Acids Research*, 36(10):3263–3273, 2008.

Li, T., Rawlings, D.J., Park, H., Kato, R.M., Witte, O.N., Satterthwaite, A.B. Constitutive membrane association potentiates activation of Bruton tyrosine kinase. *Oncogene*, 15(12):1375–1383, 1997.

Li, Y., Ye, X., Tan, C., Hongo, J.A., Zha, J., Liu, J., Kallop, D., Ludlam, M.J.C., Pei, L. Axl as a potential therapeutic target in cancer: role of Axl in tumor growth, metastasis and angiogenesis. *Oncogene*, 28(39):3442–3455, 2009.

Liao, X.C., Littman, D.R. Altered T cell receptor signaling and disrupted T cell development in mice lacking Itk. *Immunity*, 3(6):757–769, 1995.

Lichty, J.J., Malecki, J.L., Agnew, H.D., Michelson-Horowitz, D.J., Tan, S. Comparison of affinity tags for protein purification. *Protein Expression and Purification*, 41(1):98–105, 2005.

Lim, J., Hao, T., Shaw, C., Patel, A.J., Szabó, G., Rual, J.F., Fisk, C.J., Li, N., Smolyar, A., Hill, D.E., Barabási, A.L., Vidal, M., Zoghbi, H.Y. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell*, 125(4):801–814, 2006.

Lim, W.A., Pawson, T. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell*, 142(5):661–667, 2010.

Linding, R., Jensen, L.J., Ostheimer, G.J., van Vugt, M.A.T.M., Jørgensen, C., Miron, I.M., Diella, F., Colwill, K., Taylor, L., Elder, K., Metalnikov, P., Nguyen, V., Pasculescu, A., Jin, J., Park, J.G., Samson, L.D., Woodgett, J.R., Russell, R.B., Bork, P., Yaffe, M.B., Pawson, T. Systematic discovery of in vivo phosphorylation networks. *Cell*, 129(7):1415–1426, 2007.

Linger, R.M.A., Keating, A.K., Earp, H.S., Graham, D.K. Taking aim at Mer and Axl receptor tyrosine kinases as novel therapeutic targets in solid tumors. *Expert Opinion on Therapeutic Targets*, 14(10):1073–1090, 2010.

Liu, B.A., Engelmann, B.W., Jablonowski, K., Higginbotham, K., Stergachis, A.B., Nash, P.D. SRC Homology 2 Domain Binding Sites in Insulin, IGF-1 and FGF receptor mediated signaling networks reveal an extensive potential interactome. *Cell communication and signaling: CCS*, 10(1):27, 2012.

Liu, B.A., Jablonowski, K., Raina, M., Arcé, M., Pawson, T., Nash, P.D. The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. *Molecular Cell*, 22(6):851–868, 2006.

Liu, B.A., Jablonowski, K., Shah, E.E., Engelmann, B.W., Jones, R.B., Nash, P.D. SH2 domains recognize contextual peptide sequence information to determine selectivity. *Molecular & cellular proteomics: MCP*, 9(11):2391–2404, 2010.

Liu, B.A., Shah, E., Jablonowski, K., Stergachis, A., Engelmann, B., Nash, P.D. The SH2 domain-containing proteins in 21 species establish the provenance and scope of phosphotyrosine signaling in eukaryotes. *Science Signaling*, 4(202):ra83, 2011a.

Liu, J.P., Baker, J., Perkins, A.S., Robertson, E.J., Efstratiadis, A. Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type 1 IGF receptor (Igf1r). *Cell*, 75(1):59–72, 1993.

Liu, L., Makowske, M. Phosphotyrosine protein of molecular mass 30 kDa binds specifically to the positively charged region of the pleckstrin N-terminal pleckstrin homology domain. *The Biochemical Journal*, 342 (Pt 2):423–430, 1999.

Liu, M., Grigoriev, A. Protein domains correlate strongly with exons in multiple eukaryotic genomes–evidence of exon shuffling? *Trends in genetics: TIG*, 20(9):399–403, 2004.

Liu, Y.Y., Slotine, J.J., Barabási, A.L. Controllability of complex networks. *Nature*, 473(7346):167–173, 2011b.

LoPiccolo, J., Blumenthal, G.M., Bernstein, W.B., Dennis, P.A. Targeting the PI3K/Akt/mTOR pathway: effective combinations and clinical considerations. *Drug Resistance Updates: Reviews and Commentaries in Antimicrobial and Anticancer Chemotherapy*, 11(1-2):32–50, 2008.

Lowenstein, E.J., Daly, R.J., Batzer, A.G., Li, W., Margolis, B., Lammers, R., Ullrich, A., Skolnik, E.Y., Bar-Sagi, D., Schlessinger, J. The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell*, 70(3):431–442, 1992.

Lu, Q., Gore, M., Zhang, Q., Camenisch, T., Boast, S., Casagranda, F., Lai, C., Skinner, M.K., Klein, R., Matsushima, G.K., Earp, H.S., Goff, S.P., Lemke, G. Tyro-3 family receptors are essential regulators of mammalian spermatogenesis. *Nature*, 398(6729):723–728, 1999.

Lu, W., Yamamoto, V., Ortega, B., Baltimore, D. Mammalian Ryk is a Wnt coreceptor required for stimulation of neurite outgrowth. *Cell*, 119(1):97–108, 2004.

Lunter, P.C., Wiche, G. Direct binding of plectin to Fer kinase and negative regulation of its catalytic activity. *Biochemical and Biophysical Research Communications*, 296(4):904–910, 2002.

Lupher, M.L., Songyang, Z., Shoelson, S.E., Cantley, L.C., Band, H. The Cbl phosphotyrosine-binding domain selects a D(N/D)XpY motif and binds to the Tyr292 negative regulatory phosphorylation site of ZAP-70. *The Journal of Biological Chemistry*, 272(52):33140–33144, 1997.

Ma, H.W., Zeng, A.P. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics (Oxford, England)*, 19(11):1423–1430, 2003.

Machida, K., Mayer, B.J., Nollau, P. Profiling the global tyrosine phosphorylation state. *Molecular & cellular proteomics: MCP*, 2(4):215–233, 2003.

Mackarehtschian, K., Hardin, J.D., Moore, K.A., Boast, S., Goff, S.P., Lemischka, I.R. Targeted disruption of the flk2/flt3 gene leads to deficiencies in primitive hematopoietic progenitors. *Immunity*, 3(1):147–161, 1995.

Maclennan, A.J., Shaw, G. A yeast SH2 domain. *Trends in Biochemical Sciences*, 18(12):464–465, 1993.

Majeti, R., Xu, Z., Parslow, T.G., Olson, J.L., Daikh, D.I., Killeen, N., Weiss, A. An inactivating point mutation in the inhibitory wedge of CD45 causes lymphoproliferation and autoimmunity. *Cell*, 103(7):1059–1070, 2000.

Malbec, O., Fong, D.C., Turner, M., Tybulewicz, V.L., Cambier, J.C., Fridman, W.H., Daëron, M. Fc epsilon receptor I-associated lyn-dependent phosphorylation of Fc gamma receptor IIB during negative regulation of mast cell activation. *Journal of Immunology (Baltimore, Md.: 1950)*, 160(4):1647–1658, 1998.

Mallick-Wood, C.A., Pao, W., Cheng, A.M., Lewis, J.M., Kulkarni, S., Bolen, J.B., Rowley, B., Tigelaar, R.E., Pawson, T., Hayday, A.C. Disruption of epithelial gamma delta T cell repertoires by mutation of the Syk tyrosine kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 93(18):9704–9709, 1996.

Manning, G., Plowman, G.D., Hunter, T., Sudarsanam, S. Evolution of protein kinase signaling from yeast to man. *Trends in Biochemical Sciences*, 27(10):514–520, 2002a.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S. The protein kinase complement of the human genome. *Science (New York, N.Y.)*, 298(5600):1912–1934, 2002b.

Manning, G., Young, S.L., Miller, W.T., Zhai, Y. The protist, Monosiga brevicollis, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proceedings of the National Academy of Sciences of the United States of America*, 105(28):9674–9679, 2008.

Mao, X., Kikani, C.K., Riojas, R.A., Langlais, P., Wang, L., Ramos, F.J., Fang, Q., Christ-Roberts, C.Y., Hong, J.Y., Kim, R.Y., Liu, F., Dong, L.Q. APPL1 binds to adiponectin receptors and mediates adiponectin signalling and function. *Nature Cell Biology*, 8(5):516–523, 2006.

Maraskovsky, E., Daro, E., Roux, E., Teepe, M., Maliszewski, C.R., Hoek, J., Caron, D., Lebsack, M.E., McKenna, H.J. In vivo generation of human dendritic cell subsets by Flt3 ligand. *Blood*, 96(3):878–884, 2000.

Marengere, L.E., Pawson, T. Identification of residues in GTPase-activating protein Src homology 2 domains that control binding to tyrosine phosphorylated growth factor receptors and p62. *The Journal of Biological Chemistry*, 267(32):22779–22786, 1992.

Marti, F., Garcia, G.G., Lapinski, P.E., MacGregor, J.N., King, P.D. Essential role of the T cell-specific adapter protein in the activation of LCK in peripheral T cells. *The Journal of Experimental Medicine*, 203(2):281–287, 2006.

Marti, F., King, P.D. The p95-100 kDa ligand of the T cell-specific adaptor (TSAd) protein Src-homology-2 (SH2) domain implicated in TSAd nuclear import is p97 Valosin-containing protein (VCP). *Immunology Letters*, 97(2):235–243, 2005.

Marti, F., Lapinski, P.E., King, P.D. The emerging role of the T cell-specific adaptor (TSAd) protein as an autoimmune disease-regulator in mouse and man. *Immunology Letters*, 97(2):165–170, 2005.

Marti, F., Post, N.H., Chan, E., King, P.D. A transcription function for the T cell-specific adapter (TSAd) protein in T cells: critical role of the TSAd Src homology 2 domain. *The Journal of Experimental Medicine*, 193(12):1425–1430, 2001.

Marti, F., Xu, C.W., Selvakumar, A., Brent, R., Dupont, B., King, P.D. LCK-phosphorylated human killer cell-inhibitory receptors recruit and activate phosphatidylinositol 3-kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 95(20):11810–11815, 1998.

Matsumoto, T., Bohman, S., Dixelius, J., Berge, T., Dimberg, A., Magnusson, P., Wang, L., Wikner, C., Qi, J.H., Wernstedt, C., Wu, J., Bruheim, S., Mugishima, H., Mukhopadhyay, D., Spurkland, A., Claesson-Welsh, L. VEGF receptor-2 Y951 signaling and a role for the adapter molecule TSAd in tumor angiogenesis. *The EMBO journal*, 24(13):2342–2353, 2005.

Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B., Kanapin, A., Lewis, S., Mahajan, S., May, B., Schmidt, E., Vastrik, I., Wu, G., Birney, E., Stein, L., D'Eustachio, P. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Research*, 37(Database issue):D619–622, 2009.

Mauro, M.J., Druker, B.J. Chronic myelogenous leukemia. *Current Opinion in Oncology*, 13(1):3–7, 2001.

Mayer, B.J., Jackson, P.K., Van Etten, R.A., Baltimore, D. Point mutations in the abl SH2 domain coordinately impair phosphotyrosine binding in vitro and transforming activity in vivo. *Molecular and Cellular Biology*, 12(2):609–618, 1992.

Meakin, S.O., MacDonald, J.I., Gryz, E.A., Kubu, C.J., Verdi, J.M. The signaling adapter FRS-2 competes with Shc for binding to the nerve growth factor receptor TrkA. A model for discriminating proliferation and differentiation. *The Journal of Biological Chemistry*, 274(14):9861–9870, 1999.

Meerang, M., Ritz, D., Paliwal, S., Garajova, Z., Bosshard, M., Mailand, N., Janscak, P., Hübscher, U., Meyer, H., Ramadan, K. The ubiquitin-selective

segregase VCP/p97 orchestrates the response to DNA double-strand breaks. *Nature Cell Biology*, 13(11):1376–1382, 2011.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., Bork, P. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417(6887):399–403, 2002.

Miaczynska, M., Christoforidis, S., Giner, A., Shevchenko, A., Uttenweiler-Joseph, S., Habermann, B., Wilm, M., Parton, R.G., Zerial, M. APPL proteins link Rab5 to nuclear signal transduction via an endosomal compartment. *Cell*, 116(3):445–456, 2004.

Miller, M.L., Jensen, L.J., Diella, F., Jørgensen, C., Tinti, M., Li, L., Hsiung, M., Parker, S.A., Bordeaux, J., Sicheritz-Ponten, T., Olhovsky, M., Pasculescu, A., Alexander, J., Knapp, S., Blom, N., Bork, P., Li, S., Cesareni, G., Pawson, T., Turk, B.E., Yaffe, M.B., Brunak, S., Linding, R. Linear motif atlas for phosphorylation-dependent signaling. *Science Signaling*, 1(35):ra2, 2008.

Min, L., Joseph, R.E., Fulton, D.B., Andreotti, A.H. Itk tyrosine kinase substrate docking is mediated by a nonclassical SH2 domain surface of PLCgamma1. *Proceedings of the National Academy of Sciences of the United States of America*, 106(50):21143–21148, 2009.

Mitsuuchi, Y., Johnson, S.W., Sonoda, G., Tanno, S., Golemis, E.A., Testa, J.R. Identification of a chromosome 3p14.3-21.1 gene, APPL, encoding an adaptor molecule that interacts with the oncoprotein-serine/threonine kinase AKT2. *Oncogene*, 18(35):4891–4898, 1999.

Miyake, I., Ohira, M., Nakagawara, A., Sakai, R. Distinct role of ShcC docking protein in the differentiation of neuroblastoma. *Oncogene*, 28(5):662–673, 2009.

Miyoshi, K., Cui, Y., Riedlinger, G., Robinson, P., Lehoczky, J., Zon, L., Oka, T., Dewar, K., Hennighausen, L. Structure of the mouse Stat 3/5 locus: evolution from Drosophila to zebrafish to mouse. *Genomics*, 71(2):150–155, 2001.

Mizuno, K., Tagawa, Y., Mitomo, K., Watanabe, N., Katagiri, T., Ogimoto, M., Yakura, H. Src homology region 2 domain-containing phosphatase 1 positively

regulates B cell receptor-induced apoptosis by modulating association of B cell linker protein with Nck and activation of c-Jun NH2-terminal kinase. *Journal of Immunology (Baltimore, Md.: 1950)*, 169(2):778–786, 2002.

Mohamed, A.J., Yu, L., Bäckesjö, C.M., Vargas, L., Faryal, R., Aints, A., Christensson, B., Berglöf, A., Vihinen, M., Nore, B.F., Smith, C.I.E. Bruton's tyrosine kinase (Btk): function, regulation, and transformation with special emphasis on the PH domain. *Immunological Reviews*, 228(1):58–73, 2009.

Molina, T.J., Kishihara, K., Siderovski, D.P., van Ewijk, W., Narendran, A., Timms, E., Wakeham, A., Paige, C.J., Hartmann, K.U., Veillette, A. Profound block in thymocyte development in mice lacking p56lck. *Nature*, 357(6374):161–164, 1992.

Morishige, M., Hashimoto, S., Ogawa, E., Toda, Y., Kotani, H., Hirose, M., Wei, S., Hashimoto, A., Yamada, A., Yano, H., Mazaki, Y., Kodama, H., Nio, Y., Manabe, T., Wada, H., Kobayashi, H., Sabe, H. GEP100 links epidermal growth factor receptor signalling to Arf6 activation to induce breast cancer invasion. *Nature Cell Biology*, 10(1):85–92, 2008.

Mothe, I., Delahaye, L., Filloux, C., Pons, S., White, M.F., Van Obberghen, E. Interaction of wild type and dominant-negative p55PIK regulatory subunit of phosphatidylinositol 3-kinase with insulin-like growth factor-1 signaling proteins. *Molecular Endocrinology (Baltimore, Md.)*, 11(13):1911–1923, 1997.

Mousson, F., Kolkman, A., Pijnappel, W.W.M.P., Timmers, H.T.M., Heck, A.J.R. Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes. *Molecular & cellular proteomics: MCP*, 7(5):845–852, 2008.

Murphy, S.M., Bergman, M., Morgan, D.O. Suppression of c-Src activity by C-terminal Src kinase involves the c-Src SH2 and SH3 domains: analysis with Saccharomyces cerevisiae. *Molecular and Cellular Biology*, 13(9):5290–5300, 1993.

Muthusamy, N., Leiden, J.M. A protein kinase C-, Ras-, and RSK2-dependent signal transduction pathway activates the cAMP-responsive element-binding

protein transcription factor following T cell receptor engagement. *The Journal of Biological Chemistry*, 273(35):22841–22847, 1998.

Müllauer, L., Gruber, P., Sebinger, D., Buch, J., Wohlfart, S., Chott, A. Mutations in apoptosis genes: a pathogenetic factor for human disease. *Mutation Research*, 488(3):211–231, 2001.

Müller, W.E. Review: How was metazoan threshold crossed? The hypothetical Urmetazoa. *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology*, 129(2-3):433–460, 2001.

Naka, T., Fujimoto, M., Kishimoto, T. Negative regulation of cytokine signaling: STAT-induced STAT inhibitor. *Trends in Biochemical Sciences*, 24(10):394–398, 1999.

Nakagawara, A., Azar, C.G., Scavarda, N.J., Brodeur, G.M. Expression and function of TRK-B and BDNF in human neuroblastomas. *Molecular and Cellular Biology*, 14(1):759–767, 1994.

Nakao, R., Hirasaka, K., Goto, J., Ishidoh, K., Yamada, C., Ohno, A., Okumura, Y., Nonaka, I., Yasutomo, K., Baldwin, K.M., Kominami, E., Higashibata, A., Nagano, K., Tanaka, K., Yasui, N., Mills, E.M., Takeda, S., Nikawa, T. Ubiquitin ligase Cbl-b is a negative regulator for insulin-like growth factor 1 signaling during muscle atrophy caused by unloading. *Molecular and Cellular Biology*, 29(17):4798–4811, 2009.

Nakayama, M., Kikuno, R., Ohara, O. Protein-protein interactions between large proteins: two-hybrid screening using a functionally classified library composed of long cDNAs. *Genome Research*, 12(11):1773–1784, 2002.

Naramura, M., Nadeau, S., Mohapatra, B., Ahmad, G., Mukhopadhyay, C., Sattler, M., Raja, S.M., Natarajan, A., Band, V., Band, H. Mutant Cbl proteins as oncogenic drivers in myeloproliferative disorders. *Oncotarget*, 2(3):245–250, 2011.

Neel, B.G., Gu, H., Pao, L. The 'Shp'ing news: SH2 domain-containing tyrosine phosphatases in cell signaling. *Trends in Biochemical Sciences*, 28(6):284–293, 2003.

Neel, B.G., Tonks, N.K. Protein tyrosine phosphatases in signal transduction. *Current Opinion in Cell Biology*, 9(2):193–204, 1997.

Neet, K., Hunter, T. Vertebrate non-receptor protein-tyrosine kinase families. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms*, 1(2):147–169, 1996.

Negishi, I., Motoyama, N., Nakayama, K., Nakayama, K., Senju, S., Hatakeyama, S., Zhang, Q., Chan, A.C., Loh, D.Y. Essential role for ZAP-70 in both positive and negative selection of thymocytes. *Nature*, 376(6539):435–438, 1995.

Nibbe, R.K., Markowitz, S., Myeroff, L., Ewing, R., Chance, M.R. Discovery and scoring of protein interaction subnetworks discriminative of late stage human colon cancer. *Molecular & cellular proteomics: MCP*, 8(4):827–845, 2009.

Nicholson, R.I., Gee, J.M., Harper, M.E. EGFR and cancer prognosis. *European Journal of Cancer (Oxford, England: 1990)*, 37 Suppl 4:S9–15, 2001.

Nie, J., McGill, M.A., Dermer, M., Dho, S.E., Wolting, C.D., McGlade, C.J. LNX functions as a RING type E3 ubiquitin ligase that targets the cell fate determinant Numb for ubiquitin-dependent degradation. *The EMBO journal*, 21(1-2):93–102, 2002.

Nishizumi, H., Taniuchi, I., Yamanashi, Y., Kitamura, D., Ilic, D., Mori, S., Watanabe, T., Yamamoto, T. Impaired proliferation of peripheral B cells and indication of autoimmune disease in lyn-deficient mice. *Immunity*, 3(5):549–560, 1995.

Nocka, K., Tan, J.C., Chiu, E., Chu, T.Y., Ray, P., Traktman, P., Besmer, P. Molecular bases of dominant negative and loss of function mutations at the murine c-kit/white spotting locus: W37, Wv, W41 and W. *The EMBO journal*, 9(6):1805–1813, 1990.

Nomi, M., Oishi, I., Kani, S., Suzuki, H., Matsuda, T., Yoda, A., Kitamura, M., Itoh, K., Takeuchi, S., Takeda, K., Akira, S., Ikeya, M., Takada, S., Minami, Y. Loss of mRor1 enhances the heart and skeletal abnormalities in mRor2-deficient mice: redundant and pleiotropic functions of mRor1 and mRor2

receptor tyrosine kinases. *Molecular and Cellular Biology*, 21(24):8329–8335, 2001.

Normanno, N., De Luca, A., Bianco, C., Strizzi, L., Mancino, M., Maiello, M.R., Carotenuto, A., De Feo, G., Caponigro, F., Salomon, D.S. Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene*, 366(1):2–16, 2006.

Northey, J.J., Chmielecki, J., Ngan, E., Russo, C., Annis, M.G., Muller, W.J., Siegel, P.M. Signaling through ShcA is required for transforming growth factor beta- and Neu/ErbB-2-induced breast cancer cell motility and invasion. *Molecular and Cellular Biology*, 28(10):3162–3176, 2008.

O'Brien, R., Rugman, P., Renzoni, D., Layton, M., Handa, R., Hilyard, K., Waterfield, M.D., Driscoll, P.C., Ladbury, J.E. Alternative modes of binding of proteins with tandem SH2 domains. *Protein Science: A Publication of the Protein Society*, 9(3):570–579, 2000.

Oda, K., Matsuoka, Y., Funahashi, A., Kitano, H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology*, 1:2005.0010, 2005.

Ohmori, Y., Hamilton, T.A. Requirement for STAT1 in LPS-induced gene expression in macrophages. *Journal of Leukocyte Biology*, 69(4):598–604, 2001.

Okada, M., Nada, S., Yamanashi, Y., Yamamoto, T., Nakagawa, H. CSK: a protein-tyrosine kinase involved in regulation of src family kinases. *The Journal of Biological Chemistry*, 266(36):24249–24252, 1991.

Olayioye, M.A., Neve, R.M., Lane, H.A., Hynes, N.E. The ErbB signaling network: receptor heterodimerization in development and cancer. *The EMBO journal*, 19(13):3159–3167, 2000.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*, 127(3):635–648, 2006.

Ono, M., Okada, H., Bolland, S., Yanagi, S., Kurosaki, T., Ravetch, J.V. Deletion of SHIP or SHP-1 reveals two distinct pathways for inhibitory signaling. *Cell*, 90(2):293–301, 1997.

Orioli, D., Henkemeyer, M., Lemke, G., Klein, R., Pawson, T. Sek4 and Nuk receptors cooperate in guidance of commissural axons and in palate formation. *The EMBO journal*, 15(22):6035–6049, 1996.

Osborne, M.A., Dalton, S., Kochan, J.P. The yeast tribrid system–genetic detection of trans-phosphorylated ITAM-SH2-interactions. *Bio/Technology (Nature Publishing Company)*, 13(13):1474–1478, 1995.

O'Shea, J.J. Jaks, STATs, cytokine signal transduction, and immunoregulation: are we there yet? *Immunity*, 7(1):1–11, 1997.

Ostman, A., Böhmer, F.D. Regulation of receptor tyrosine kinase signaling by protein tyrosine phosphatases. *Trends in Cell Biology*, 11(6):258–266, 2001.

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J.i., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T.o., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T.,

Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T., Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nature Genetics*, 36(1):40–45, 2004.

Oyama, M., Kozuka-Hata, H., Tasaki, S., Semba, K., Hattori, S., Sugano, S., Inoue, J.i., Yamamoto, T. Temporal perturbation of tyrosine phosphoproteome dynamics reveals the system-wide regulatory networks. *Molecular & cellular proteomics: MCP*, 8(2):226–231, 2009.

Palidwor, G.A., Shcherbinin, S., Huska, M.R., Rasko, T., Stelzl, U., Arumughan, A., Foulle, R., Porras, P., Sanchez-Pulido, L., Wanker, E.E., Andrade-Navarro, M.A. Detection of alpha-rod protein repeats using a neural network and application to huntingtin. *PLoS computational biology*, 5(3):e1000304, 2009.

Pandey, A., Ibarrola, N., Kratchmarova, I., Fernandez, M.M., Constantinescu, S.N., Ohara, O., Sawasdikosol, S., Lodish, H.F., Mann, M. A novel Src homology 2 domain-containing molecule, Src-like adapter protein-2 (SLAP-2), which negatively regulates T cell receptor signaling. *The Journal of Biological Chemistry*, 277(21):19131–19138, 2002.

Pappu, R., Cheng, A.M., Li, B., Gong, Q., Chiu, C., Griffin, N., White, M., Sleckman, B.P., Chan, A.C. Requirement for B cell linker protein (BLNK) in B cell development. *Science (New York, N.Y.)*, 286(5446):1949–1954, 1999.

Park, D., Choi, Y.B., Han, M.K., Kim, U.H., Shin, J., Yun, Y. Adaptor protein Lad relays PDGF signal to Grb2 in lung cells: a tissue-specific PDGF signal transduction. *Biochemical and Biophysical Research Communications*, 284(2):275–281, 2001.

Park, D., Yun, Y. Tyrosine phosphorylation-dependent yeast two-hybrid system for the identification of the SH2 domain-binding proteins. *Molecules and Cells*, 12(2):244–249, 2001.

Park, S., Frisén, J., Barbacid, M. Aberrant axonal projections in mice lacking EphA8 (Eek) tyrosine protein kinase receptors. *The EMBO journal*, 16(11):3106–3114, 1997.

Parrish, J.R., Gulyas, K.D., Finley, R.L. Yeast two-hybrid contributions to interactome mapping. *Current Opinion in Biotechnology*, 17(4):387–393, 2006.

Patsialou, A., Wilsker, D., Moran, E. DNA-binding properties of ARID family proteins. *Nucleic Acids Research*, 33(1):66–80, 2005.

Paul, W.E. *Fundamental Immunology*, volume 6. Lippincott Williams & Wilkins, 2008.

Pawson, T., Gish, G.D., Nash, P. SH2 domains, interaction modules and cellular wiring. *Trends in Cell Biology*, 11(12):504–511, 2001.

Pawson, T., Gish, G.D., Nash, P. The SH2 domain: a prototype for protein interaction modules. *Protein Science Encyclopedia*, 2005.

Pawson, T., Scott, J.D. Signaling through scaffold, anchoring, and adaptor proteins. *Science (New York, N.Y.)*, 278(5346):2075–2080, 1997.

Perchonock, C.E., Fernando, M.C., Quinn, W.J., Nguyen, C.T., Sun, J., Shapiro, M.J., Shapiro, V.S. Negative regulation of interleukin-2 and p38 mitogen-activated protein kinase during T-cell activation by the adaptor ALX. *Molecular and Cellular Biology*, 26(16):6005–6015, 2006.

Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18):3551–3567, 1999.

Pfeffer, L.M., Mullersman, J.E., Pfeffer, S.R., Murti, A., Shi, W., Yang, C.H. STAT3 as an adapter to couple phosphatidylinositol 3-kinase to the IFNAR1 chain of the type I interferon receptor. *Science (New York, N.Y.)*, 276(5317):1418–1420, 1997.

Pflieger, D., Jünger, M.A., Müller, M., Rinner, O., Lee, H., Gehrig, P.M., Gstaiger, M., Aebersold, R. Quantitative proteomic analysis of protein complexes:

concurrent identification of interactors and their state of phosphorylation. *Molecular & cellular proteomics: MCP*, 7(2):326–346, 2008.

Phizicky, E., Bastiaens, P.I.H., Zhu, H., Snyder, M., Fields, S. Protein analysis on a proteomic scale. *Nature*, 422(6928):208–215, 2003.

Pincus, D., Letunic, I., Bork, P., Lim, W.A. Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages. *Proceedings of the National Academy of Sciences of the United States of America*, 105(28):9680–9684, 2008.

Pivniouk, V.I., Martin, T.R., Lu-Kuo, J.M., Katz, H.R., Oettgen, H.C., Geha, R.S. SLP-76 deficiency impairs signaling via the high-affinity IgE receptor in mast cells. *The Journal of Clinical Investigation*, 103(12):1737–1743, 1999.

Pixley, F.J., Stanley, E.R. CSF-1 regulation of the wandering macrophage: complexity in action. *Trends in Cell Biology*, 14(11):628–638, 2004.

Pons, S., Asano, T., Glasheen, E., Miralpeix, M., Zhang, Y., Fisher, T.L., Myers, M.G., Sun, X.J., White, M.F. The structure and function of p55PIK reveal a new regulatory subunit for phosphatidylinositol 3-kinase. *Molecular and Cellular Biology*, 15(8):4453–4465, 1995.

Pujana, M.A., Han, J.D.J., Starita, L.M., Stevens, K.N., Tewari, M., Ahn, J.S., Rennert, G., Moreno, V., Kirchhoff, T., Gold, B., Assmann, V., Elshamy, W.M., Rual, J.F., Levine, D., Rozek, L.S., Gelman, R.S., Gunsalus, K.C., Greenberg, R.A., Sobhian, B., Bertin, N., Venkatesan, K., Ayivi-Guedehoussou, N., Solé, X., Hernández, P., Lázaro, C., Nathanson, K.L., Weber, B.L., Cusick, M.E., Hill, D.E., Offit, K., Livingston, D.M., Gruber, S.B., Parvin, J.D., Vidal, M. Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nature Genetics*, 39(11):1338–1349, 2007.

Puri, M.C., Rossant, J., Alitalo, K., Bernstein, A., Partanen, J. The receptor tyrosine kinase TIE is required for integrity and survival of vascular endothelial cells. *The EMBO journal*, 14(23):5884–5891, 1995.

Radke, K., Gilmore, T., Martin, G.S. Transformation by Rous sarcoma virus: a cellular substrate for transformation-specific protein phosphorylation contains phosphotyrosine. *Cell*, 21(3):821–828, 1980.

Radtke, S., Haan, S., Jörissen, A., Hermanns, H.M., Diefenbach, S., Smyczek, T., Schmitz-Vandeleur, H., Heinrich, P.C., Behrmann, I., Haan, C. The Jak1 SH2 domain does not fulfill a classical SH2 function in Jak/STAT signaling but plays a structural role for receptor interaction and up-regulation of receptor surface expression. *The Journal of Biological Chemistry*, 280(27):25760–25768, 2005.

Rain, J.C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schächter, V., Chemama, Y., Labigne, A., Legrain, P. The protein-protein interaction map of Helicobacter pylori. *Nature*, 409(6817):211–215, 2001.

Rajagopal, K., Sommers, C.L., Decker, D.C., Mitchell, E.O., Korthauer, U., Sperling, A.I., Kozak, C.A., Love, P.E., Bluestone, J.A. RIBP, a novel Rlk/Txk- and itk-binding adaptor protein that regulates T cell activation. *The Journal of Experimental Medicine*, 190(11):1657–1668, 1999.

Rameh, L.E., Arvidsson, A.k., Carraway, K.L., Couvillon, A.D., Rathbun, G., Crompton, A., VanRenterghem, B., Czech, M.P., Ravichandran, K.S., Burakoff, S.J., Wang, D.S., Chen, C.S., Cantley, L.C. A comparative analysis of the phosphoinositide binding specificity of pleckstrin homology domains. *The Journal of Biological Chemistry*, 272(35):22059–22066, 1997.

Rameh, L.E., Chen, C.S., Cantley, L.C. Phosphatidylinositol (3,4,5)P3 interacts with SH2 domains and modulates PI 3-kinase association with tyrosine-phosphorylated proteins. *Cell*, 83(5):821–830, 1995.

Ramponi, G., Stefani, M. Structural, catalytic, and functional properties of low M(r), phosphotyrosine protein phosphatases. Evidence of a long evolutionary history. *The International Journal of Biochemistry & Cell Biology*, 29(2):279–292, 1997.

Rashid, T., Upton, A.L., Blentic, A., Ciossek, T., Knöll, B., Thompson, I.D., Drescher, U. Opposing gradients of ephrin-As and EphA7 in the superior colliculus are essential for topographic mapping in the mammalian visual system. *Neuron*, 47(1):57–69, 2005.

Ravichandran, K.S., Zhou, M.M., Pratt, J.C., Harlan, J.E., Walk, S.F., Fesik, S.W., Burakoff, S.J. Evidence for a requirement for both phospholipid and phosphotyrosine binding via the Shc phosphotyrosine-binding domain in vivo. *Molecular and Cellular Biology*, 17(9):5540–5549, 1997.

Rebecchi, M.J., Scarlata, S. Pleckstrin homology domains: a common fold with diverse functions. *Annual Review of Biophysics and Biomolecular Structure*, 27:503–528, 1998.

Reed, A.L., Yamazaki, H., Kaufman, J.D., Rubinstein, Y., Murphy, B., Johnson, A.C. Molecular cloning and characterization of a transcription regulator with homology to GC-binding factor. *The Journal of Biological Chemistry*, 273(34):21594–21602, 1998.

Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B.J., Hon, G.C., Myers, C.L., Parsons, A., Friesen, H., Oughtred, R., Tong, A., Stark, C., Ho, Y., Botstein, D., Andrews, B., Boone, C., Troyanskya, O.G., Ideker, T., Dolinski, K., Batada, N.N., Tyers, M. Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. *Journal of Biology*, 5(4):11, 2006.

Remy, I., Montmarquette, A., Michnick, S.W. PKB/Akt modulates TGF-beta signalling through a direct interaction with Smad3. *Nature Cell Biology*, 6(4):358–365, 2004.

Ren, R., Ye, Z.S., Baltimore, D. Abl protein-tyrosine kinase selects the Crk adapter as a substrate using SH3-binding sites. *Genes & Development*, 8(7):783–795, 1994.

Resnik, P. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th international joint conference on Artificial intelligence*, volume 1, pp. 448–453. Morgan Kaufmann Publishers Inc., 1995.

Rhee, S.G., Chang, T.S., Bae, Y.S., Lee, S.R., Kang, S.W. Cellular regulation by hydrogen peroxide. *Journal of the American Society of Nephrology: JASN*, 14(8 Suppl 3):S211–215, 2003.

Rice, D.S., Northcutt, G.M., Kurschner, C. The Lnx family proteins function as molecular scaffolds for Numb family proteins. *Molecular and Cellular Neurosciences*, 18(5):525–540, 2001.

Riethmacher, D., Sonnenberg-Riethmacher, E., Brinkmann, V., Yamaai, T., Lewin, G.R., Birchmeier, C. Severe neuropathies in mice with targeted mutations in the ErbB3 receptor. *Nature*, 389(6652):725–730, 1997.

Rinner, O., Mueller, L.N., Hubálek, M., Müller, M., Gstaiger, M., Aebersold, R. An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. *Nature Biotechnology*, 25(3):345–352, 2007.

Rizo, J., Südhof, T.C. C2-domains, structure and function of a universal Ca2+-binding domain. *The Journal of Biological Chemistry*, 273(26):15879–15882, 1998.

Robinson, D.R., Wu, Y.M., Lin, S.F. The protein tyrosine kinase family of the human genome. *Oncogene*, 19(49):5548–5557, 2000.

Rocchi, S., Tartare-Deckert, S., Murdaca, J., Holgado-Madruga, M., Wong, A.J., Van Obberghen, E. Determination of Gab1 (Grb2-associated binder-1) interaction with insulin receptor-signaling molecules. *Molecular Endocrinology (Baltimore, Md.)*, 12(7):914–923, 1998.

Roll, J.D., Reuther, G.W. ALK-activating homologous mutations in LTK induce cellular transformation. *PloS One*, 7(2):e31733, 2012.

Rual, J.F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P.O., Clingingsmith, T.R., Hartley, J.L., Esposito, D., Cheo, D., Moore, T., Simmons, B., Sequerra, R., Bosak, S., Doucette-Stamm, L., Le Peuch, C., Vandenhaute, J., Cusick, M.E., Albala, J.S., Hill, D.E., Vidal, M. Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Research*, 14(10B):2128–2135, 2004.

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D.S., Zhang, L.V.,

Wong, S.L., Franklin, G., Li, S., Albala, J.S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R.S., Vandenhaute, J., Zoghbi, H.Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M.E., Hill, D.E., Roth, F.P., Vidal, M. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437(7062):1173–1178, 2005.

Rubin, B.P., Singer, S., Tsao, C., Duensing, A., Lux, M.L., Ruiz, R., Hibbard, M.K., Chen, C.J., Xiao, S., Tuveson, D.A., Demetri, G.D., Fletcher, C.D., Fletcher, J.A. KIT activation is a ubiquitous feature of gastrointestinal stromal tumors. *Cancer Research*, 61(22):8118–8121, 2001.

Rui, L., Yuan, M., Frantz, D., Shoelson, S., White, M.F. SOCS-1 and SOCS-3 block insulin signaling by ubiquitin-mediated degradation of IRS1 and IRS2. *The Journal of Biological Chemistry*, 277(44):42394–42398, 2002.

Rusch, V., Baselga, J., Cordon-Cardo, C., Orazem, J., Zaman, M., Hoda, S., McIntosh, J., Kurie, J., Dmitrovsky, E. Differential expression of the epidermal growth factor receptor and its ligands in primary non-small cell lung cancers and adjacent benign lung. *Cancer Research*, 53(10 Suppl):2379–2385, 1993.

Rush, J., Moritz, A., Lee, K.A., Guo, A., Goss, V.L., Spek, E.J., Zhang, H., Zha, X.M., Polakiewicz, R.D., Comb, M.J. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nature Biotechnology*, 23(1):94–101, 2005.

Sadowski, I., Stone, J.C., Pawson, T. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Molecular and Cellular Biology*, 6(12):4396–4408, 1986.

Salim, K., Bottomley, M.J., Querfurth, E., Zvelebil, M.J., Gout, I., Scaife, R., Margolis, R.L., Gigg, R., Smith, C.I., Driscoll, P.C., Waterfield, M.D., Panayotou, G. Distinct specificity in the recognition of phosphoinositides by the pleckstrin homology domains of dynamin and Bruton's tyrosine kinase. *The EMBO journal*, 15(22):6241–6250, 1996.

Sandlund, J.T., Shurtleff, S.A., Onciu, M., Horwitz, E., Leung, W., Howard, V., Rencher, R., Conley, M.E. Frequent mutations in SH2D1A (XLP) in

males presenting with high-grade mature B-cell neoplasms. *Pediatric Blood & Cancer*, 60(9):E85–87, 2013.

Santoro, M., Carlomagno, F., Melillo, R.M., Fusco, A. Dysfunction of the RET receptor in human cancer. *Cellular and molecular life sciences: CMLS*, 61(23):2954–2964, 2004.

Sattler, M., Salgia, R., Shrikhande, G., Verma, S., Pisick, E., Prasad, K.V., Griffin, J.D. Steel factor induces tyrosine phosphorylation of CRKL and binding of CRKL to a complex containing c-kit, phosphatidylinositol 3-kinase, and p120(CBL). *The Journal of Biological Chemistry*, 272(15):10248–10253, 1997.

Savelieva, K.V., Rajan, I., Baker, K.B., Vogel, P., Jarman, W., Allen, M., Lanthorn, T.H. Learning and memory impairment in Eph receptor A6 knockout mice. *Neuroscience Letters*, 438(2):205–209, 2008.

Sayós, J., Martínez-Barriocanal, A., Kitzig, F., Bellón, T., López-Botet, M. Recruitment of C-terminal Src kinase by the leukocyte inhibitory receptor CD85j. *Biochemical and Biophysical Research Communications*, 324(2):640–647, 2004.

Schad, E., Tompa, P., Hegyi, H. The relationship between proteome size, structural disorder and organism complexity. *Genome Biology*, 12(12):R120, 2011.

Schaefer, C.F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., Buetow, K.H. PID: the Pathway Interaction Database. *Nucleic Acids Research*, 37(Database issue):D674–679, 2009.

Schlessinger, J., Ullrich, A. Growth factor signaling by receptor tyrosine kinases. *Neuron*, 9(3):383–391, 1992.

Scholz, G., Cartledge, K., Dunn, A.R. Hck enhances the adherence of lipopolysaccharide-stimulated macrophages via Cbl and phosphatidylinositol 3-kinase. *The Journal of Biological Chemistry*, 275(19):14615–14623, 2000.

Schuchardt, A., D'Agati, V., Larsson-Blomberg, L., Costantini, F., Pachnis, V. Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. *Nature*, 367(6461):380–383, 1994.

Schulze, W.X., Deng, L., Mann, M. Phosphotyrosine interactome of the ErbB-receptor kinase family. *Molecular Systems Biology*, 1:2005.0008, 2005.

Schulze, W.X., Mann, M. A novel proteomic screen for peptide-protein interactions. *The Journal of Biological Chemistry*, 279(11):10756–10764, 2004.

Scott, J., Ideker, T., Karp, R.M., Sharan, R. Efficient algorithms for detecting signaling pathways in protein interaction networks. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 13(2):133–144, 2006.

Sefton, B.M., Hunter, T., Beemon, K., Eckhart, W. Evidence that the phosphorylation of tyrosine is essential for cellular transformation by Rous sarcoma virus. *Cell*, 20(3):807–816, 1980.

Serfling, E., Avots, A., Neumann, M. The architecture of the interleukin-2 promoter: a reflection of T lymphocyte activation. *Biochimica Et Biophysica Acta*, 1263(3):181–200, 1995.

Shalaby, F., Rossant, J., Yamaguchi, T.P., Gertsenstein, M., Wu, X.F., Breitman, M.L., Schuh, A.C. Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature*, 376(6535):62–66, 1995.

Shaywitz, A.J., Dove, S.L., Kornhauser, J.M., Hochschild, A., Greenberg, M.E. Magnitude of the CREB-dependent transcriptional response is determined by the strength of the interaction between the kinase-inducible domain of CREB and the KIX domain of CREB-binding protein. *Molecular and Cellular Biology*, 20(24):9409–9422, 2000.

Shimoda, K., van Deursen, J., Sangster, M.Y., Sarawar, S.R., Carson, R.T., Tripp, R.A., Chu, C., Quelle, F.W., Nosaka, T., Vignali, D.A., Doherty, P.C., Grosveld, G., Paul, W.E., Ihle, J.N. Lack of IL-4-induced Th2 response and IgE class switching in mice with disrupted Stat6 gene. *Nature*, 380(6575):630–633, 1996.

Shimoyama, M., Matsuoka, H., Nagata, A., Iwata, N., Tamekane, A., Okamura, A., Gomyo, H., Ito, M., Jishage, K.i., Kamada, N., Suzuki, H., Tetsuo Noda, T., Matsui, T. Developmental expression of EphB6 in the thymus: lessons from EphB6 knockout mice. *Biochemical and Biophysical Research Communications*, 298(1):87–94, 2002.

Simister, P.C., Burton, N.M., Brady, R.L. Phosphotyrosine recognition by the Raf kinase inhibitor protein. In *Forum on Immunopathological Diseases and Therapeutics*, volume 2. Begel House Inc., 2011.

Simonis, N., Rual, J.F., Carvunis, A.R., Tasan, M., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Sahalie, J.M., Venkatesan, K., Gebreab, F., Cevik, S., Klitgord, N., Fan, C., Braun, P., Li, N., Ayivi-Guedehoussou, N., Dann, E., Bertin, N., Szeto, D., Dricot, A., Yildirim, M.A., Lin, C., de Smet, A.S., Kao, H.L., Simon, C., Smolyar, A., Ahn, J.S., Tewari, M., Boxem, M., Milstein, S., Yu, H., Dreze, M., Vandenhaute, J., Gunsalus, K.C., Cusick, M.E., Hill, D.E., Tavernier, J., Roth, F.P., Vidal, M. Empirically controlled mapping of the Caenorhabditis elegans protein-protein interactome network. *Nature Methods*, 6(1):47–54, 2009.

Siraganian, R.P., Zhang, J., Suzuki, K., Sada, K. Protein tyrosine kinase Syk in mast cell signaling. *Molecular Immunology*, 38(16-18):1229–1233, 2002.

Sitko, J.C., Guevara, C.I., Cacalano, N.A. Tyrosine-phosphorylated SOCS3 interacts with the Nck and Crk-L adapter proteins and regulates Nck activation. *The Journal of Biological Chemistry*, 279(36):37662–37669, 2004.

Skålhegg, B.S., Taskén, K., Hansson, V., Huitfeldt, H.S., Jahnsen, T., Lea, T. Location of cAMP-dependent protein kinase type I with the TCR-CD3 complex. *Science (New York, N.Y.)*, 263(5143):84–87, 1994.

Smart, J.E., Oppermann, H., Czernilofsky, A.P., Purchio, A.F., Erikson, R.L., Bishop, J.M. Characterization of sites for tyrosine phosphorylation in the transforming protein of Rous sarcoma virus (pp60v-src) and its normal cellular homologue (pp60c-src). *Proceedings of the National Academy of Sciences of the United States of America*, 78(10):6013–6017, 1981.

Smend, S., Nienstädt, F., Grossmann, A. Method for changing modes in an electronic device. 2014. US Patent App. 13/844,459.

Smeyne, R.J., Klein, R., Schnapp, A., Long, L.K., Bryant, S., Lewin, A., Lira, S.A., Barbacid, M. Severe sensory and sympathetic neuropathies in mice carrying a disrupted Trk/NGF receptor gene. *Nature*, 368(6468):246–249, 1994.

Smith, J.M., Szathmary, E. *The major transitions in evolution*. Oxford University Press, 1997.

Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R.J. SH2 domains recognize specific phosphopeptide sequences. *Cell*, 72(5):767–778, 1993.

Songyang, Z., Shoelson, S.E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X.R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T. Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Molecular and Cellular Biology*, 14(4):2777–2785, 1994.

Sonnenberg-Riethmacher, E., Walter, B., Riethmacher, D., Gödecke, S., Birchmeier, C. The c-ros tyrosine kinase receptor controls regionalization and differentiation of epithelial cells in the epididymis. *Genes & Development*, 10(10):1184–1193, 1996.

Soriano, P. Abnormal kidney development and hematological disorders in PDGF beta-receptor mutant mice. *Genes & Development*, 8(16):1888–1896, 1994.

Soriano, P. The PDGF alpha receptor is required for neural crest cell development and for normal patterning of the somites. *Development (Cambridge, England)*, 124(14):2691–2700, 1997.

Spirin, V., Mirny, L.A. Protein complexes and functional modules in molecular networks. *Proceedings of the National Academy of Sciences of the United States of America*, 100(21):12123–12128, 2003.

Stahelin, R.V., Kong, K.F., Raha, S., Tian, W., Melowic, H.R., Ward, K.E., Murray, D., Altman, A., Cho, W. Protein kinase Cϑ C2 domain is a phosphotyrosine binding module that plays a key role in its activation. *The Journal of Biological Chemistry*, 287(36):30518–30528, 2012.

Starr, R., Hilton, D.J. SOCS: suppressors of cytokine signalling. *The International Journal of Biochemistry & Cell Biology*, 30(10):1081–1085, 1998.

Steelman, L.S., Abrams, S.L., Whelan, J., Bertrand, F.E., Ludwig, D.E., Bäsecke, J., Libra, M., Stivala, F., Milella, M., Tafuri, A., Lunghi, P., Bonati, A., Martelli, A.M., McCubrey, J.A. Contributions of the Raf/MEK/ERK, PI3K/PTEN/Akt/mTOR and Jak/STAT pathways to leukemia. *Leukemia*, 22(4):686–707, 2008.

Stelzl, U., Wanker, E.E. The value of high quality protein-protein interaction networks for systems biology. *Current Opinion in Chemical Biology*, 10(6):551–558, 2006.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksöz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., Wanker, E.E. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 122(6):957–968, 2005.

Stephen, L.J., Fawkes, A.L., Verhoeve, A., Lemke, G., Brown, A. A critical role for the EphA3 receptor tyrosine kinase in heart development. *Developmental Biology*, 302(1):66–79, 2007.

Stotz, A., Linder, P. The ADE2 gene from Saccharomyces cerevisiae: sequence and new vectors. *Gene*, 95(1):91–98, 1990.

Streubel, B., Vinatzer, U., Willheim, M., Raderer, M., Chott, A. Novel t(5;9)(q33;q22) fuses ITK to SYK in unspecified peripheral T-cell lymphoma. *Leukemia*, 20(2):313–318, 2006.

Stumpf, M.P.H., Thorne, T., de Silva, E., Stewart, R., An, H.J., Lappe, M., Wiuf, C. Estimating the size of the human interactome. *Proceedings of the National*

Academy of Sciences of the United States of America, 105(19):6959–6964, 2008.

Suga, H., Torruella, G., Burger, G., Brown, M.W., Ruiz-Trillo, I. Earliest holozoan expansion of phosphotyrosine signaling. *Molecular Biology and Evolution*, 31(3):517–528, 2014.

Sun, Z., Li, X., Massena, S., Kutschera, S., Padhan, N., Gualandi, L., Sundvold-Gjerstad, V., Gustafsson, K., Choy, W.W., Zang, G., Quach, M., Jansson, L., Phillipson, M., Abid, M.R., Spurkland, A., Claesson-Welsh, L. VEGFR2 induces c-Src signaling and vascular permeability in vivo via the adaptor protein TSAd. *The Journal of Experimental Medicine*, 209(7):1363–1377, 2012.

Suter, B., Fontaine, J.F., Yildirimman, R., Raskó, T., Schaefer, M.H., Rasche, A., Porras, P., Vázquez-Álvarez, B.M., Russ, J., Rau, K., Foulle, R., Zenkner, M., Saar, K., Herwig, R., Andrade-Navarro, M.A., Wanker, E.E. Development and application of a DNA microarray-based yeast two-hybrid system. *Nucleic Acids Research*, 41(3):1496–1507, 2013.

Sylvester, M., Kliche, S., Lange, S., Geithner, S., Klemm, C., Schlosser, A., Grossmann, A., Stelzl, U., Schraven, B., Krause, E., Freund, C. Adhesion and degranulation promoting adapter protein (ADAP) is a central hub for phosphotyrosine-mediated interactions in T cells. *PloS One*, 5(7):e11708, 2010.

Szathmáry, E., Smith, J.M. The major evolutionary transitions. *Nature*, 374(6519):227–232, 1995.

Takaki, S., Sauer, K., Iritani, B.M., Chien, S., Ebihara, Y., Tsuji, K., Takatsu, K., Perlmutter, R.M. Control of B cell production by the adaptor protein lnk. Definition Of a conserved family of signal-modulating proteins. *Immunity*, 13(5):599–609, 2000.

Takeda, K., Kaisho, T., Yoshida, N., Takeda, J., Kishimoto, T., Akira, S. Stat3 activation is responsible for IL-6-dependent T cell proliferation through

preventing apoptosis: generation and characterization of T cell-specific Stat3-deficient mice. *Journal of Immunology (Baltimore, Md.: 1950)*, 161(9):4652–4660, 1998.

Takeda, K., Noguchi, K., Shi, W., Tanaka, T., Matsumoto, M., Yoshida, N., Kishimoto, T., Akira, S. Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proceedings of the National Academy of Sciences of the United States of America*, 94(8):3801–3804, 1997.

Takeuchi, S., Takeda, K., Oishi, I., Nomi, M., Ikeya, M., Itoh, K., Tamura, S., Ueda, T., Hatta, T., Otani, H., Terashima, T., Takada, S., Yamamura, H., Akira, S., Minami, Y. Mouse Ror2 receptor tyrosine kinase is required for the heart development and limb formation. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms*, 5(1):71–78, 2000.

Tan, C.S.H., Bodenmiller, B., Pasculescu, A., Jovanovic, M., Hengartner, M.O., Jørgensen, C., Bader, G.D., Aebersold, R., Pawson, T., Linding, R. Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Science Signaling*, 2(81):ra39, 2009a.

Tan, C.S.H., Pasculescu, A., Lim, W.A., Pawson, T., Bader, G.D., Linding, R. Positive selection of tyrosine loss in metazoan evolution. *Science (New York, N.Y.)*, 325(5948):1686–1688, 2009b.

Tan, P.B., Kim, S.K. Signaling specificity: the RTK/RAS/MAP kinase pathway in metazoans. *Trends in genetics: TIG*, 15(4):145–149, 1999.

Tangye, S.G., Phillips, J.H., Lanier, L.L., Nichols, K.E. Functional requirement for SAP in 2B4-mediated activation of human natural killer cells as revealed by the X-linked lymphoproliferative syndrome. *Journal of Immunology (Baltimore, Md.: 1950)*, 165(6):2932–2936, 2000.

Taraghi, B., Saranti, A., Ebner, M., Grossmann, A., Müller, V. Adaptive Learner Profiling Provides the Optimal Sequence of Posed Basic Mathematical Problems. In C. Rensing, S. de Freitas, T. Ley, P. Muñoz-Merino, eds., *Open Learning and Teaching in Educational Communities*, volume 8719 of *Lecture Notes in Computer Science*, pp. 592–593. Springer International Publishing, 2014a.

Taraghi, Behnam Frey, M., Saranti, A., Ebner, M., Müller, V., Grossmann, A. Determining the Causing Factors of Errors for Multiplication Problems. In *Proceedings of the 2014 Immersive Education Summit*, pp. 144–153. Immersive Education Initiative, 2014b.

Tarasenko, T., Kole, H.K., Chi, A.W., Mentink-Kane, M.M., Wynn, T.A., Bolland, S. T cell-specific deletion of the inositol phosphatase SHIP reveals its role in regulating Th1/Th2 and cytotoxic responses. *Proceedings of the National Academy of Sciences of the United States of America*, 104(27):11382–11387, 2007.

Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., Bull, S., Pawson, T., Morris, Q., Wrana, J.L. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnology*, 27(2):199–204, 2009.

Temple, G., Lamesch, P., Milstein, S., Hill, D.E., Wagner, L., Moore, T., Vidal, M. From genome to proteome: developing expression clone resources for the human genome. *Human Molecular Genetics*, 15 Spec No 1:R31–43, 2006.

Threadgill, D.W., Dlugosz, A.A., Hansen, L.A., Tennenbaum, T., Lichti, U., Yee, D., LaMantia, C., Mourton, T., Herrup, K., Harris, R.C. Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science (New York, N.Y.)*, 269(5221):230–234, 1995.

Tonks, N.K. Protein tyrosine phosphatases: from genes, to function, to disease. *Nature Reviews. Molecular Cell Biology*, 7(11):833–846, 2006.

Trivier, E., Ganesan, T.S. RYK, a catalytically inactive receptor tyrosine kinase, associates with EphB2 and EphB3 but does not interact with AF-6. *The Journal of Biological Chemistry*, 277(25):23037–23043, 2002.

Turner, M., Mee, P.J., Walters, A.E., Quinn, M.E., Mellor, A.L., Zamoyska, R., Tybulewicz, V.L. A requirement for the Rho-family GTP exchange factor Vav in positive and negative selection of thymocytes. *Immunity*, 7(4):451–460, 1997.

Turner, N., Grose, R. Fibroblast growth factor signalling: from development to cancer. *Nature Reviews. Cancer*, 10(2):116–129, 2010.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., Rothberg, J.M. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, 403(6770):623–627, 2000.

Uhlik, M.T., Temple, B., Bencharit, S., Kimple, A.J., Siderovski, D.P., Johnson, G.L. Structural and evolutionary division of phosphotyrosine binding (PTB) domains. *Journal of Molecular Biology*, 345(1):1–20, 2005.

Ullrich, A., Schlessinger, J. Signal transduction by receptors with tyrosine kinase activity. *Cell*, 61(2):203–212, 1990.

Vang, T., Torgersen, K.M., Sundvold, V., Saxena, M., Levy, F.O., Skålhegg, B.S., Hansson, V., Mustelin, T., Taskén, K. Activation of the COOH-terminal Src kinase (Csk) by cAMP-dependent protein kinase inhibits signaling through the T cell receptor. *The Journal of Experimental Medicine*, 193(4):497–507, 2001.

Venkatesan, K., Rual, J.F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.I., Yildirim, M.A., Simonis, N., Heinzmann, K., Gebreab, F., Sahalie, J.M., Cevik, S., Simon, C., de Smet, A.S., Dann, E., Smolyar, A., Vinayagam, A., Yu, H., Szeto, D., Borick, H., Dricot, A., Klitgord, N., Murray, R.R., Lin, C., Lalowski, M., Timm, J., Rau, K., Boone, C., Braun, P., Cusick, M.E., Roth, F.P., Hill, D.E., Tavernier, J., Wanker, E.E., Barabási, A.L., Vidal, M. An empirical framework for binary interactome mapping. *Nature Methods*, 6(1):83–90, 2009.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan,

M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T.,

Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X. The sequence of the human genome. *Science (New York, N.Y.)*, 291(5507):1304–1351, 2001.

Verbeek, B.S., Adriaansen-Slot, S.S., Rijksen, G., Vroom, T.M. Grb2 overexpression in nuclei and cytoplasm of human breast cells: a histochemical and biochemical study of normal and neoplastic mammary tissue specimens. *The Journal of Pathology*, 183(2):195–203, 1997.

Vinayagam, A., Stelzl, U., Foulle, R., Plassmann, S., Zenkner, M., Timm, J., Assmus, H.E., Andrade-Navarro, M.A., Wanker, E.E. A directed protein interaction network for investigating intracellular signal transduction. *Science Signaling*, 4(189):rs8, 2011.

Vogel, C., Chothia, C. Protein family expansions and biological complexity. *PLoS computational biology*, 2(5):e48, 2006.

Vogel, W.F., Aszódi, A., Alves, F., Pawson, T. Discoidin domain receptor 1 tyrosine kinase has an essential role in mammary gland development. *Molecular and Cellular Biology*, 21(8):2906–2917, 2001.

Vogelstein, B., Kinzler, K.W. Cancer genes and the pathways they control. *Nature Medicine*, 10(8):789–799, 2004.

Volkmer, R., Tapia, V., Landgraf, C. Synthetic peptide arrays for investigating protein interaction domains. *FEBS letters*, 586(17):2780–2786, 2012.

Vossler, M.R., Yao, H., York, R.D., Pan, M.G., Rim, C.S., Stork, P.J. cAMP activates MAP kinase and Elk-1 through a B-Raf- and Rap1-dependent pathway. *Cell*, 89(1):73–82, 1997.

Wade, N. A decade later, genetic map yields few new cures. *New York Times*, 12, 2010.

Waksman, G., Kominos, D., Robertson, S.C., Pant, N., Baltimore, D., Birge, R.B., Cowburn, D., Hanafusa, H., Mayer, B.J., Overduin, M., Resh, M.D., Rios, C.B., Silverman, L., Kuriyan, J. Crystal structure of the phosphotyrosine

recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature*, 358(6388):646–653, 1992.

Walhout, A.J., Vidal, M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods (San Diego, Calif.)*, 24(3):297–306, 2001.

Waltz, S.E., Eaton, L., Toney-Earley, K., Hess, K.A., Peace, B.E., Ihlendorf, J.R., Wang, M.H., Kaestner, K.H., Degen, S.J. Ron-mediated cytoplasmic signaling is dispensable for viability but is required to limit inflammatory responses. *The Journal of Clinical Investigation*, 108(4):567–576, 2001.

Wang, D., Feng, J., Wen, R., Marine, J.C., Sangster, M.Y., Parganas, E., Hoffmeyer, A., Jackson, C.W., Cleveland, J.L., Murray, P.J., Ihle, J.N. Phospholipase Cgamma2 is essential in the functions of B cell and several Fc receptors. *Immunity*, 13(1):25–35, 2000.

Wang, X., Huang, L. Identifying dynamic interactors of protein complexes by quantitative mass spectrometry. *Molecular & cellular proteomics: MCP*, 7(1):46–57, 2008.

Warner, A.J., Lopez-Dee, J., Knight, E.L., Feramisco, J.R., Prigent, S.A. The Shc-related adaptor protein, Sck, forms a complex with the vascular-endothelial-growth-factor receptor KDR in transfected cells. *The Biochemical Journal*, 347(Pt 2):501–509, 2000.

Warren, M., Luthra, R., Yin, C.C., Ravandi, F., Cortes, J.E., Kantarjian, H.M., Medeiros, L.J., Zuo, Z. Clinical impact of change of FLT3 mutation status in acute myeloid leukemia patients. *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc*, 25(10):1405–1412, 2012.

Wavreille, A.S., Garaud, M., Zhang, Y., Pei, D. Defining SH2 domain and PTP specificity by screening combinatorial peptide libraries. *Methods (San Diego, Calif.)*, 42(3):207–219, 2007.

Weimann, M., Grossmann, A., Woodsmith, J., Özkan, Z., Birth, P., Meierhofer, D., Benlasfer, N., Valovka, T., Timmermann, B., Wanker, E.E., Sauer, S.,

Stelzl, U. A Y2H-seq approach defines the human protein methyltransferase interactome. *Nature Methods*, 10(4):339–342, 2013.

Weinstein, M., Xu, X., Ohyama, K., Deng, C.X. FGFR-3 and FGFR-4 function cooperatively to direct alveogenesis in the murine lung. *Development (Cambridge, England)*, 125(18):3615–3623, 1998.

Weisberg, H.I. *Bias and causation: Models and judgment for valid comparisons*, volume 885. John Wiley & Sons, 2010.

Wen, S.T., Van Etten, R.A. The PAG gene product, a stress-induced protein with antioxidant properties, is an Abl SH3-binding protein and a physiological inhibitor of c-Abl tyrosine kinase activity. *Genes & Development*, 11(19):2456–2467, 1997.

Wertheimer, E., Lu, S.P., Backeljauw, P.F., Davenport, M.L., Taylor, S.I. Homozygous deletion of the human insulin receptor gene results in leprechaunism. *Nature Genetics*, 5(1):71–73, 1993.

Wesoly, J., Agarwal, S., Sigurdsson, S., Bussen, W., Van Komen, S., Qin, J., van Steeg, H., van Benthem, J., Wassenaar, E., Baarends, W.M., Ghazvini, M., Tafel, A.A., Heath, H., Galjart, N., Essers, J., Grootegoed, J.A., Arnheim, N., Bezzubova, O., Buerstedde, J.M., Sung, P., Kanaar, R. Differential contributions of mammalian Rad54 paralogs to recombination, DNA damage repair, and meiosis. *Molecular and Cellular Biology*, 26(3):976–989, 2006.

White, M.F. Insulin signaling in health and disease. *Science (New York, N.Y.)*, 302(5651):1710–1711, 2003.

Williams, S.E., Mann, F., Erskine, L., Sakurai, T., Wei, S., Rossi, D.J., Gale, N.W., Holt, C.E., Mason, C.A., Henkemeyer, M. Ephrin-B2 and EphB1 mediate retinal axon divergence at the optic chiasm. *Neuron*, 39(6):919–935, 2003.

Wilson, A.L., Schrecengost, R.S., Guerrero, M.S., Thomas, K.S., Bouton, A.H. Breast cancer antiestrogen resistance 3 (BCAR3) promotes cell motility by regulating actin cytoskeletal and adhesion remodeling in invasive breast cancer cells. *PloS One*, 8(6):e65678, 2013.

Wilson, E.B. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927.

Wojcik, J., Schächter, V. Protein-protein interaction map inference using interacting domain profile pairs. *Bioinformatics (Oxford, England)*, 17 Suppl 1:S296–305, 2001.

Wolf, G., Trüb, T., Ottinger, E., Groninga, L., Lynch, A., White, M.F., Miyazaki, M., Lee, J., Shoelson, S.E. PTB domains of IRS-1 and Shc have distinct but overlapping binding specificities. *The Journal of Biological Chemistry*, 270(46):27407–27410, 1995.

Wong, J., Nakajima, Y., Westermann, S., Shang, C., Kang, J.S., Goodner, C., Houshmand, P., Fields, S., Chan, C.S.M., Drubin, D., Barnes, G., Hazbun, T. A protein interaction map of the mitotic spindle. *Molecular Biology of the Cell*, 18(10):3800–3809, 2007.

Woodsmith, J., Kamburov, A., Stelzl, U. Dual coordination of post translational modifications in human protein networks. *PLoS computational biology*, 9(3):e1002933, 2013.

Worseck, J.M., Grossmann, A., Weimann, M., Hegele, A., Stelzl, U. A stringent yeast two-hybrid matrix screening approach for protein-protein interaction discovery. *Methods in Molecular Biology (Clifton, N.J.)*, 812:63–87, 2012.

Wu, L.W., Mayo, L.D., Dunbar, J.D., Kessler, K.M., Ozes, O.N., Warren, R.S., Donner, D.B. VRAP is an adaptor protein that binds KDR, a receptor for vascular endothelial cell growth factor. *The Journal of Biological Chemistry*, 275(9):6059–6062, 2000.

Wu, S.L., Kim, J., Bandle, R.W., Liotta, L., Petricoin, E., Karger, B.L. Dynamic profiling of the post-translational modifications and interaction partners of epidermal growth factor receptor signaling after stimulation by epidermal growth factor using Extended Range Proteomic Analysis (ERPA). *Molecular & cellular proteomics: MCP*, 5(9):1610–1627, 2006.

Xu, X., Sarikas, A., Dias-Santagata, D.C., Dolios, G., Lafontant, P.J., Tsai, S.C., Zhu, W., Nakajima, H., Nakajima, H.O., Field, L.J., Wang, R., Pan,

Z.Q. The CUL7 E3 ubiquitin ligase targets insulin receptor substrate 1 for ubiquitin-dependent degradation. *Molecular Cell*, 30(4):403–414, 2008.

Yaffe, M.B. Phosphotyrosine-binding domains in signal transduction. *Nature Reviews. Molecular Cell Biology*, 3(3):177–186, 2002.

Yamada, K., Kawata, H., Shou, Z., Hirano, S., Mizutani, T., Yazawa, T., Sekiguchi, T., Yoshino, M., Kajitani, T., Miyamoto, K. Analysis of zinc-fingers and homeoboxes (ZHX)-1-interacting proteins: molecular cloning and characterization of a member of the ZHX family, ZHX3. *The Biochemical Journal*, 373(Pt 1):167–178, 2003.

Yamada, M., Suzuki, K., Mizutani, M., Asada, A., Matozaki, T., Ikeuchi, T., Koizumi, S., Hatanaka, H. Analysis of tyrosine phosphorylation-dependent protein-protein interactions in TrkB-mediated intracellular signaling using modified yeast two-hybrid system. *Journal of Biochemistry*, 130(1):157–165, 2001.

Yamamoto, S., Sakai, N., Nakamura, H., Fukagawa, H., Fukuda, K., Takagi, T. INOH: ontology-based highly structured database of signal transduction pathways. *Database: The Journal of Biological Databases and Curation*, 2011:bar052, 2011.

Yang, X.J. Multisite protein modification and intramolecular signaling. *Oncogene*, 24(10):1653–1662, 2005.

Yaoi, T., Chamnongpol, S., Jiang, X., Li, X. Src homology 2 domain-based high throughput assays for profiling downstream molecules in receptor tyrosine kinase pathways. *Molecular & cellular proteomics: MCP*, 5(5):959–968, 2006.

Yarden, Y., Sliwkowski, M.X. Untangling the ErbB signalling network. *Nature Reviews. Molecular Cell Biology*, 2(2):127–137, 2001.

Yook, S.H., Oltvai, Z.N., Barabási, A.L. Functional and topological characterization of protein interaction networks. *Proteomics*, 4(4):928–942, 2004.

Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., Wang, S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics (Oxford, England)*, 26(7):976–978, 2010.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., Hao, T., Rual, J.F., Dricot, A., Vazquez, A., Murray, R.R., Simon, C., Tardivo, L., Tam, S., Svrzikapa, N., Fan, C., de Smet, A.S., Motyl, A., Hudson, M.E., Park, J., Xin, X., Cusick, M.E., Moore, T., Boone, C., Snyder, M., Roth, F.P., Barabási, A.L., Tavernier, J., Hill, D.E., Vidal, M. High-quality binary protein interaction map of the yeast interactome network. *Science (New York, N.Y.)*, 322(5898):104–110, 2008.

Yu, J.W., Lemmon, M.A. Genome-wide analysis of signaling domain function. *Current Opinion in Chemical Biology*, 7(1):103–109, 2003.

Yun, M., Keshvara, L., Park, C.G., Zhang, Y.M., Dickerson, J.B., Zheng, J., Rock, C.O., Curran, T., Park, H.W. Crystal structures of the Dab homology domains of mouse disabled 1 and 2. *The Journal of Biological Chemistry*, 278(38):36572–36581, 2003.

Zhang, J.G., Farley, A., Nicholson, S.E., Willson, T.A., Zugaro, L.M., Simpson, R.J., Moritz, R.L., Cary, D., Richardson, R., Hausmann, G., Kile, B.J., Kent, S.B., Alexander, W.S., Metcalf, D., Hilton, D.J., Nicola, N.A., Baca, M. The conserved SOCS box motif in suppressors of cytokine signaling binds to elongins B and C and may couple bound proteins to proteasomal degradation. *Proceedings of the National Academy of Sciences of the United States of America*, 96(5):2071–2076, 1999.

Zhang, X., Ibrahimi, O.A., Olsen, S.K., Umemori, H., Mohammadi, M., Ornitz, D.M. Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family. *The Journal of Biological Chemistry*, 281(23):15694–15700, 2006.

Zhang, Z.Y. Protein-tyrosine phosphatases: biological function, structural characteristics, and mechanism of catalysis. *Critical Reviews in Biochemistry and Molecular Biology*, 33(1):1–52, 1998.

Zhao, B., Tan, P.H., Li, S.S.C., Pei, D. Systematic characterization of the specificity of the SH2 domains of cytoplasmic tyrosine kinases. *Journal of Proteomics*, 81:56–69, 2013.

Zheng, D., Gu, S., Li, Y., Ji, C., Xie, Y., Mao, Y. A global genomic view on LNX siRNA-mediated cell cycle arrest. *Molecular Biology Reports*, 38(4):2771–2783, 2011.

Zhong, J., Yang, P., Muta, K., Dong, R., Marrero, M., Gong, F., Wang, C.Y. Loss of Jak2 selectively suppresses DC-mediated innate immune response and protects mice from lethal dose of LPS-induced septic shock. *PloS One*, 5(3):e9593, 2010.

Zhou, M.M., Huang, B., Olejniczak, E.T., Meadows, R.P., Shuker, S.B., Miyazaki, M., Trüb, T., Shoelson, S.E., Fesik, S.W. Structural basis for IL-4 receptor phosphopeptide recognition by the IRS-1 PTB domain. *Nature Structural Biology*, 3(4):388–393, 1996.

Zhou, M.M., Ravichandran, K.S., Olejniczak, E.F., Petros, A.M., Meadows, R.P., Sattler, M., Harlan, J.E., Wade, W.S., Burakoff, S.J., Fesik, S.W. Structure and ligand recognition of the phosphotyrosine binding domain of Shc. *Nature*, 378(6557):584–592, 1995.

Zwahlen, C., Li, S.C., Kay, L.E., Pawson, T., Forman-Kay, J.D. Multiple modes of peptide recognition by the PTB domain of the cell fate determinant Numb. *The EMBO journal*, 19(7):1505–1515, 2000.

Zwick, E., Hackel, P.O., Prenzel, N., Ullrich, A. The EGF receptor as central transducer of heterologous signalling systems. *Trends in Pharmacological Sciences*, 20(10):408–412, 1999.

Appendix A

Interacting ORFs

For each interacting ORF, i.e. each ORF involved in a successful experiment, the best matching RefSeqRNA identifier is provided, including the matching region and the quality of the match.
In a post-hoc analysis, the ORF nucleotide sequences were matched against the RefSeqRNA sequences (Release 67) using BLAST, as described for Release 31. For better clarity, only one RefSeq identifier is provided for each ORF, provided matching regions are identical. The criteria applied to choose among RefSeq identifiers matching equally were, in order of priority, (1) absolute length of coding sequence (CDS) covered, percentage of CDS covered, known or predicted (NM_ or XM_), percentage of RefSeq sequence covered, matching primary human genome assembly. Where all of the criteria failed, the oldest identifier, i.e. the one with the lowest number, was chosen. ORFs containing phosphotyrosine-recognizing domains are shown in boldface. St = Start; Len = Length.
Note that, due to the ongoing annotation efforts, certain RefSeq matches differ between release 31 and release 67 with regard to Entrez GeneID. The respective RefSeq identifiers have been marked (†).

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
ABL1	NM_007313.2	99	1	3450	3450	440	3889	5881
ABL2	NM_005158.4	99	1	3504	3504	46	3549	11957
ACAP1	NM_014716.3	100	1	2223	2223	207	2429	2523
ADCK5	XM_006716527.1	100	1	510	510	1297	1806	1974
AHDC1	NM_001029882.3	100	1	1200	1200	4444	5643	6374
AKT1	NM_001014431.1	100	1	1441	1443	341	1781	2794
ANAPC10	NM_014885.4	100	1	558	558	102	659	1457
ANGPT1	NM_001199859.1	100	1	442	444	1231	1672	4335

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
ANKRD50	NM_001167882.1	100	1	2232	2232	1822	4053	7519
ANKS1A	NM_015245.2	99	1	1381	1383	2165	3545	6355
ANKS1B	NM_001204066.1	99	1	1074	1074	135	1208	3063
ANKZF1	NM_001282792.1	100	17	774	881	1392	2149	2193
ANLN	NM_001284302.1	100	1	1218	1218	2265	3482	4687
AP1S2	NM_003916.4	100	1	474	474	328	801	2320
APBB3	NM_133173.2	99	1	1460	1461	360	1819	2146
APPL1								
ORF #1	NM_012096.2	99	1	2130	2130	148	2277	6061
ORF #2	NM_012096.2	99	1	2129	2130	148	2276	6061
APPL2	NM_018171.3	99	1	1995	1995	219	2213	3289
ARAP3	XM_005268499.1	100	1	225	225	3948	4172	5200
ARFGAP1	NM_001281482.1	100	1	1212	1212	141	1352	3524
ARHGAP10	XM_005263216.2	99	13	656	660	1566	2209	2910
ARID5A	XM_005263860.1	100	1	1293	1293	428	1720	2041
ARL6IP4	NM_001002252.2	100	1	648	648	698	1345	1681
ASB3	NM_016115.4	100	1	1557	1557	203	1759	2290
ASB8	NM_024095.3	100	1	867	867	170	1036	2629
ASB9	NM_001031739.2	100	1	885	885	298	1182	1714
ASCL4	NM_203436.2	100	1	519	519	835	1353	2260
ATF3	NM_001030287.3	100	1	546	546	155	700	1935
BAIAP2								
ORF #1	NM_001144888.1	100	1	1536	1539	109	1644	3306
ORF #2	XM_005256948.1	99	1	1569	1569	112	1680	3198
BAT2L	NM_013318.3	100	1	975	1018	56	1030	11062
BATF3	NM_018664.2	100	1	384	384	224	607	992
BCAR3	NM_003567.3	100	1	2478	2478	253	2730	3203
BECN1	NM_003766.3	100	1	1353	1353	163	1515	2143
BEX5	NM_001012978.2	100	1	336	336	261	596	840
BLK	NM_001715.2	99	1	1518	1518	582	2099	2642

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
BMX								
ORF #1	NM_001721.6	100	1	2026	2028	112	2137	2530
ORF #2	NM_001721.6	100	1	2028	2028	112	2139	2530
ORF #3	NM_001721.6	100	1	2028	2028	112	2139	2530
BRWD1	NM_001007246.2	100	1	363	363	296	658	2653
BTK								
ORF #1	NM_000061.2	100	1	1980	1980	194	2173	2611
ORF #2	NM_000061.2	99	1	1978	1980	194	2171	2611
C10orf18	NM_017782.4	99	1	2226	2226	5693	7918	8636
C10orf81	XM_005270162.2	99	1	849	849	1025	1873	2172
C17orf53	XM_006722075.1	99	1	1224	1566	92	1315	1964
C17orf53	XM_006722075.1	100	1223	1566	1566	1539	1882	1964
C17orf82	NM_203425.1	100	1	756	756	226	981	1546
C19orf66	NM_018381.2	100	1	489	492	299	787	2139
C1orf135	NM_024037.2	100	1	1074	1074	102	1175	2178
C1orf62	NM_152763.4	99	1	1245	1248	280	1524	3063
C1QTNF3	NM_030945.3	99	1	381	381	502	882	3624
C21orf77	NM_144659.5	99	1	381	381	1697	2077	2730
C22orf28	NM_014306.4	99	1	1518	1518	132	1649	2081
C22orf29	NM_024627.5	100	1	912	912	695	1606	6693
C22orf39	NM_001166242.1	99	1	243	243	545	787	1558
C3orf10	NM_018462.4	100	1	228	228	54	281	1197
C3orf34	NM_032898.4	100	1	492	492	437	928	2216
C4orf17	NM_032149.2	99	1	1184	1184	363	1546	1652
C5orf35	NM_153706.3	99	1	900	900	387	1286	1722
C6orf125	NM_032340.3	100	1	381	381	66	446	1355
C6orf141	NM_001145652.1	100	1	714	714	415	1128	1696
C6orf146	NM_173563.2	99	1	1527	1527	407	1933	2270
C8orf33	NM_023080.2	100	1	690	690	55	744	2617
C9orf43	NM_001278629.1	100	1	1386	1386	408	1793	2091

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
CA8	NM_004056.4	99	1	417	462	249	665	2278
CAMK1	NM_003656.4	100	1	1113	1113	186	1298	1500
CAPN10	NM_023083.3	99	1	2019	2019	197	2215	2662
CBL	NM_005188.3	100	1	2721	2721	143	2863	11241
CBLB								
ORF #1	NM_170662.3	99	1	2949	2949	323	3271	3976
ORF #2	NM_170662.3	99	1	2949	2949	323	3271	3976
CBLC	NM_012116.3	99	1	1423	1425	64	1486	1591
CCDC14	XM_005247714.2	99	1	2262	2262	345	2606	3665
CCDC33	NM_182791.3	99	1	1026	1026	50	1077	2141
CCDC74A	NM_001258306.1	100	1	939	939	138	1076	1357
CCDC87	NM_018219.2	99	1	2550	2550	69	2618	2915
CCM2	NM_031443.3	100	1	1335	1335	147	1481	1904
CDCA4	NM_145701.2	99	1	726	726	97	822	1927
CENPB	NM_001810.5	97	29	860	899	1604	2434	2856
CHN2	NM_004067.3	99	1	1406	1407	438	1843	3461
CHTF18	NM_022092.2	99	1	2928	2928	64	2991	3096
CKM	NM_001824.4	99	1	1146	1146	176	1321	1666
CNDP2	NM_018235.2	100	1	1428	1428	260	1687	5089
COQ9	NM_020312.3	100	1	957	957	82	1038	1699
CORO6	XM_005258056.2	100	1	714	714	301	1014	1764
COX6B2	NM_144613.4	100	1	267	267	184	450	1679
CRBN	NM_001173482.1	99	1	1326	1326	31	1356	2593
CRK								
ORF #1	NM_016823.3	99	1	915	915	151	1065	3225
ORF #2	NM_005206.4	100	1	615	615	151	765	3055
ORF #3	NM_016823.3	99	1	913	915	151	1063	3225
ORF #4	NM_016823.3	99	1	915	915	151	1065	3225
CRKL								
ORF #1	NM_005207.3	100	1	910	912	510	1419	5336

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
ORF #2	NM_005207.3	100	1	910	912	510	1419	5336
CRLF3	NM_015986.3	100	1	1329	1329	110	1438	2954
CRP	NM_000567.2	100	1	141	141	639	779	2024
CRYBA2	NM_057094.1	100	1	594	594	36	629	709
CSAD	NM_001244706.1	99	16	943	943	587	1513	1757
CSK	NM_001127190.1	100	1	1350	1350	184	1533	2236
CSRP3	NM_003476.4	99	1	585	585	240	824	1464
CTNNBL1	NM_001281495.1	100	1	936	936	1026	1961	2089
CYB561D2	NM_001291284.1	100	1	669	669	316	984	1348
CYP2C8	NM_000770.3	99	1	1473	1473	96	1568	1924
CYP46A1	NM_006668.1	100	1	1503	1503	1	1503	2138
DAPP1								
ORF #1	NM_014395.2	100	1	843	843	91	933	2953
ORF #2	NM_014395.2	100	1	799	799	76	874	2953
DBN1	NM_004395.3	99	1	1950	1950	173	2122	2942
DDIT4L	NM_145244.3	100	1	582	582	246	827	2649
DDX5	NM_004396.3	100	17	608	608	186	777	3769
DOK1	NM_001381.3	100	1	1446	1446	371	1816	2269
DOK2								
ORF #1	NM_003974.2	99	1	1239	1239	94	1332	1870
ORF #2	NM_003974.2	99	1	1239	1239	94	1332	1870
DOK3								
ORF #1	NM_001144875.1	99	1	993	993	203	1195	2354
ORF #2	NM_001144875.1	99	1	993	993	203	1195	2354
DOK4	NM_018110.3	100	1	981	981	351	1331	2750
DOK5	NM_177959.2	100	1	597	597	446	1042	1734
DOK7	NM_173660.4	99	1	1515	1515	71	1585	2583
DTNA	NM_001128175.1	100	1	1116	1116	348	1463	1728
DUSP18	NM_152511.3	100	1	165	354	502	666	2453
DUSP18	NM_152511.3	100	164	354	354	775	965	2453

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
E2F2	NM_004091.3	100	1	252	252	4109	4360	5201
E2F6	NM_198256.3	100	1	846	846	299	1144	3246
EFHC1	NM_018100.3	99	1	1923	1923	216	2138	5596
ELK1	NM_005229.4	100	1	1287	1287	217	1503	2828
EPB41L4A	XM_005272043.2	100	1	555	558	452	1006	4797
EPS8	NM_004447.5	99	1	2469	2469	438	2906	4088
EPS8L2	NM_022772.3	99	1	2148	2148	248	2395	3156
EPYC								
ORF #1	NM_004950.4	99	1	969	969	94	1062	1553
ORF #2	NM_004950.4	99	1	969	969	94	1062	1553
ESD	NM_001984.1	99	1	849	849	184	1032	1208
EVI1	XM_006713537.1	100	1	375	510	397	771	2994
EYA3	NM_001990.3	100	1	1083	1089	183	1265	6040
EZH2	NM_004456.4	100	1	2256	2256	194	2449	2723
FABP7	NM_001446.3	100	1	399	399	295	693	1005
FAM114A1	NM_138389.2	99	1	1692	1692	260	1951	4138
FAM117B	NM_173511.3	100	1	1038	1038	743	1780	5795
FAM127C	NM_001078173.1	100	1	342	342	78	419	2038
FAM46A	NM_017633.2	99	1	136	1344	319	454	5617
FAM46A	NM_017633.2	100	85	1344	1344	388	1647	5617
FAM46B	NM_052943.3	100	1	1275	1275	176	1450	2391
FAM59A	NM_001242409.1	100	1	2631	2631	57	2687	7051
FANCG	NM_004629.1	100	1	1869	1869	493	2361	2649
FER	NM_005246.2	100	35	2501	2530	385	2851	2950
FES	NM_002005.3	99	1	2469	2469	97	2565	2783
FGF12	NM_004113.5	99	1	546	546	247	792	5408
FGF13	NM_004114.3	100	1	738	738	663	1400	2705
FGF21	NM_019113.2	99	1	630	630	151	780	940
FGR	NM_001042729.1	99	1	1589	1590	213	1801	2442
FOXO1	NM_007062.1†	99	1	1504	1506	88	1591	1853

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
FRK	NM_002031.2	100	1	1518	1518	448	1965	2864
FRS2	NM_001278357.1	99	1	1536	1539	304	1839	6676
FRS3								
ORF #1	NM_006653.4	100	1	1479	1479	258	1736	2198
ORF #2	NM_006653.4	100	1	1475	1479	258	1732	2198
FSTL1	NM_007085.4	100	1	927	927	176	1102	3840
FYB	NM_199335.3	100	1	891	891	1625	2515	4720
FYN								
ORF #1	NM_002037.5	100	1	1613	1614	608	2220	3628
ORF #2	NM_153048.3	100	1	1449	1449	104	1552	2959
SH2	NM_002037.5	100	34	379	397	1031	1376	3628
GABPB2								
ORF #1	NM_144618.2	100	1	1347	1347	332	1678	1953
ORF #2	NM_144618.2	100	1	1347	1347	332	1678	1953
GALNT3	XM_006712402.1	100	1	362	426	532	893	1680
GATA1	NM_002049.3	100	1	869	1008	92	960	1501
GBA3	NM_001128432.2	100	1	489	489	103	591	1268
GBL	XM_005255480.2	99	1	981	981	328	1308	1990
GLIS3	NM_152629.3	99	1	2328	2328	195	2522	6672
GRAP2								
ORF #1	NM_004810.3	100	1	993	993	264	1256	3514
ORF #2	NM_004810.3	100	1	993	993	264	1256	3514
GRB2								
ORF #1	NM_002086.4	100	1	654	654	358	1011	3303
ORF #2	NM_002086.4	99	1	651	651	358	1008	3303
SH2	NM_002086.4	100	1	309	309	529	837	3303
GRB7	NM_001030002.2	100	1	1599	1599	121	1719	2130
GRB10	NM_001001550.2	99	1	1609	1611	286	1894	4793
GRB14	NM_004490.2	99	1	1623	1623	542	2164	2402
GSTCD	NM_001031720.3	99	1	1899	1899	73	1974	4122

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
HABP4	XM_005251812.1	96	16	826	849	160	963	2210
HCK	NM_001172129.1	100	1	1518	1518	301	1818	2168
HESX1	NM_003865.2	100	1	558	558	335	892	1182
HLA-C	NM_002117.5	96	1	1101	1101	66	1166	1586
HSH2D								
ORF #1	NM_032855.3	100	1	1059	1059	532	1590	2403
ORF #2	NM_032855.3	100	1	1059	1059	532	1590	2403
HSPA1A								
ORF #1	NM_153201.3†	99	17	594	595	309	886	2014
ORF #2	NM_005346.4†	99	1	1926	1926	217	2142	2551
HSPD1								
ORF #1	NM_199440.1	98	17	852	922	142	982	2319
ORF #1	NM_199440.1	96	886	912	922	1021	1048	2319
ORF #2	NM_199440.1	100	1	1722	1722	118	1839	2319
HUWE1	XM_005261972.2	99	18	945	965	11826	12753	13271
ID1	NM_002165.3	100	1	468	468	106	573	1000
ID2	NM_002166.4	100	1	405	405	184	588	1402
IFT140	NM_014714.3	99	16	959	960	2528	3471	5277
ING4	NM_001127582.1	100	1	750	750	42	791	1461
INO80E	NM_173618.1	100	1	735	735	102	836	1185
IQUB	XM_005250162.2	100	1	1119	1122	150	1268	2487
IRS1	NM_005544.2	100	1	3729	3729	53	3781	8743
ISL1	NM_002202.2	99	1	1050	1050	549	1598	2729
ITGB1BP1								
ORF #1	NM_004763.3	99	1	603	603	178	780	1930
ORF #2	NM_004763.3	99	1	599	603	178	776	1930
ITK	NM_005546.3	100	1	1863	1863	83	1945	4366
JAK2	NM_004972.3	99	1	3397	3399	495	3891	5285
JAK3	NM_000215.3	99	1	1787	1860	101	1887	5449
KCTD13	NM_178863.4	99	1	990	990	200	1189	1745

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
KCTD17	NM_001282684.1	99	1	945	945	26	970	1779
KCTD4	NM_198404.2	99	1	780	780	405	1184	2133
KIAA0317	NM_001039479.1	99	1	2369	2370	506	2874	5445
KLF15	NM_014079.3	100	1	1251	1251	183	1433	2539
KLHL20	NM_014458.3	100	1	599	693	180	778	3505
KLRAQ1	NM_001193475.1	100	1	1088	1092	186	1273	3055
KRIT1	NM_194454.1	99	1	2209	2211	546	2754	4523
KRTAP10-7	NM_198689.2	100	1	1113	1113	26	1138	1606
KRTAP23-1	NM_181624.1	100	1	198	198	1	198	208
KRTAP3-1	NM_031958.1	100	1	297	297	41	337	614
KRTCAP2	NM_173852.3	100	1	489	489	27	515	577
LASP1	NM_006148.3	99	1	599	972	332	930	4135
LCK								
ORF #1	NM_001042771.2	100	1	964	1620	122	1085	2102
ORF #1	NM_001042771.2	100	1054	1620	1620	1085	1651	2102
ORF #2	NM_001042771.2	100	1	964	1530	122	1085	2102
ORF #2	NM_001042771.2	100	1054	1530	1530	1085	1561	2102
LCP2	NM_005565.3	100	1	1602	1602	208	1809	2472
LDHAL6B	NM_033195.2	99	1	1146	1146	126	1271	1773
LECT1	NM_001011705.1	100	1	1002	1002	112	1113	1535
LEFTY2	NM_003240.3	100	1	1101	1101	244	1344	2187
LETMD1	NM_015416.4	100	1	1083	1083	59	1141	2151
LGALS9C	NM_001042685.1†	99	1	1071	1071	66	1133	1243
LHX8	NM_001001933.1	99	1	1071	1071	665	1735	2393
LIX1	NM_153234.4	99	1	849	849	241	1089	3979
LMX1A	XM_006711319.1	99	1	303	303	700	1002	1013
LNX1	NM_001126328.2	100	1	2187	2187	286	2472	3199
LNX2	NM_153371.3	100	1	2073	2073	310	2382	4804
LOC285398	NM_052859.3	100	1	375	375	3210	3584	5112
LOC492311	NM_001007189.1	100	1	144	144	2560	2703	2900

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
LRRFIP1	NM_001137551.1	100	1	318	318	1160	1477	3599
LSM8	NM_016200.4	100	1	289	291	193	481	12520
LYN								
ORF #1	NM_002350.3	99	1	1539	1539	283	1821	4158
ORF #2	NM_002350.3	99	1	1539	1539	283	1821	4158
MAGEC3	NM_177456.2	100	1	1041	1041	528	1568	1707
MAPK8IP2	NM_012324.4	99	1	2475	2475	135	2609	3395
MAPKAPK2	NM_032960.3	99	17	917	917	1078	1979	2997
MAPRE3	NM_012326.2	100	1	846	846	154	999	1880
MATK								
ORF #1	NM_139355.2	100	1	1524	1524	401	1924	2163
ORF #2	NM_139355.2	100	1	1522	1524	401	1922	2163
MIST	NM_052964.2	100	1	1286	1287	138	1423	1710
MLL3	NM_170606.2	99	17	976	976	326	1288	16872
MLL4	XM_006723515.1	97	15	683	936	2453	3120	3907
MNDA	NM_002432.1	100	1	1224	1224	201	1424	1670
MPEG1	NM_001039396.1	99	1	2151	2151	157	2307	4527
MPG	NM_001015052.2	100	1	882	882	141	1022	1096
MPP5	NM_022474.3	99	1	2026	2028	476	2501	5608
MPZL1	NM_003953.5	100	1	444	444	3334	3777	5026
MRE11A	NM_005590.3	100	1	618	621	297	914	5164
MRPS22	NM_020191.2	99	1	1083	1083	9	1091	1155
MSRB3	NM_001193461.1	100	1	558	558	137	694	4296
MST4	NM_001042452.1	100	1	339	414	11	349	2875
MST4	NM_001042452.1	100	335	414	414	796	875	2875
MT2A	NM_005953.3	100	1	186	186	91	276	466
MTF2	NM_001164392.1	100	1	1611	1611	293	1903	3973
MYBPHL	NM_001010985.2	99	1	1065	1065	51	1115	1372
MYH7B	NM_020884.4	100	1	524	528	513	1036	6602
MYLIP	NM_013262.3	100	1	1338	1338	238	1575	3086

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
MYOZ1	NM_021245.3	100	1	900	900	366	1265	1583
MYOZ2	NM_016599.4	99	1	795	795	214	1008	2604
MZF1	NM_001267033.1	99	1	873	873	562	1434	2694
NACAD	NM_001146334.1	99	29	826	827	3255	4052	4778
NARG1L	NM_001110798.1	99	1	1290	1290	325	1614	1833
NCK1								
ORF #1	NM_006153.5	100	1	1134	1134	131	1264	4421
ORF #2	NM_006153.5	100	1	1134	1134	131	1264	4421
NCK2	NM_001004720.2	100	1	1141	1143	167	1307	2417
NGEF	NM_019850.2	99	1	2133	2133	249	2381	3184
NHLRC2	NM_198514.3	100	1	1104	1104	1319	2422	6409
NHP2	NM_017838.3	100	1	462	462	144	605	867
NKAP								
ORF #1	NM_024528.3	99	1	1248	1248	168	1415	1600
ORF #2	NM_024528.3	100	1	1248	1248	168	1415	1600
NME4	NM_005009.2	100	1	564	564	32	595	1059
NMU	NM_006681.3	99	1	525	525	107	631	834
NOS1AP	NM_014697.2	99	1	1521	1521	388	1908	4464
NT5C	XM_006721851.1	99	1	276	354	83	358	1136
NT5C	XM_006721851.1	100	276	354	354	382	460	1136
NUFIP2	NM_020772.2	100	1	2088	2088	90	2177	10897
NUMB								
ORF #1	NM_001005745.1	100	1	622	738	321	942	3470
ORF #2	XM_006720296.1	100	1	408	408	1031	1438	2811
NXT2	NM_018698.4	99	1	594	594	103	696	2713
OBSL1	NM_001173408.1	98	1	186	186	2884	3069	3578
OFCC1	XM_003118558.3	100	1	343	696	387	729	1595
OLIG1	NM_138983.2	100	1	768	768	152	919	2293
OSAP	NM_032623.3	100	1	723	723	181	903	1339
OSBPL6	XM_005246268.1	99	1	1527	1527	541	2067	2232

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
OTUD7B	NM_020205.3	99	1	1283	1284	357	1639	6415
P4HA2	NM_001017974.1	100	1	1515	1515	180	1694	2110
PACRGL	NM_145048.3	100	1	666	666	392	1057	2090
PACSIN3	NM_001184974.1	100	1	1275	1275	211	1485	1913
PAFAH1B2	NM_002572.3	100	1	690	690	143	832	4200
PAQR7	NM_178422.5	99	1	1041	1041	668	1708	3023
PARD6A	NM_001037281.1	100	1	1038	1038	92	1129	1270
PCDHB5	NM_015669.3	99	1	2388	2388	230	2617	2926
PDPK1								
ORF #1	NM_031268.5	99	1	628	638	812	1439	6862
ORF #2	NM_031268.5	99	1	449	457	989	1437	6862
PELI1	NM_020651.3	99	1	1257	1257	461	1717	3780
PELI3	NM_145065.2	99	1	842	996	165	1006	2763
PELO	NM_015946.4	99	1	1158	1158	986	2143	2941
PER1	XM_005256690.1	100	1	855	885	239	1093	4529
PGRMC1								
ORF #1	NM_006667.4	98	17	610	610	92	683	1931
ORF #2	NM_006667.4	100	1	588	588	120	707	1931
PHC2	NM_004427.3	99	1	972	972	353	1324	2566
PIK3CA	NM_006218.2	99	1	3205	3207	158	3362	3724
PIK3CB	NM_006219.2	100	1	3212	3213	17	3228	5931
PIK3R1	NM_181504.3	99	1	1365	1365	91	1455	5691
PIK3R2	NM_005027.3	99	1	762	762	1966	2727	3981
PIK3R3								
ORF #1	NM_001114172.1	99	1	1386	1386	258	1643	5194
ORF #2	NM_001114172.1	99	1	1385	1386	258	1642	5194
ORF #3	NM_001114172.1	99	1	1385	1386	258	1642	5194
ORF #4	NM_001114172.1	99	1	1386	1386	258	1643	5194
ORF #5	NM_001114172.1	99	1	1386	1386	258	1643	5194
PINK1	NM_032409.2	99	1	1746	1746	95	1840	2680

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
PLAGL2	NM_002657.3	100	1	1491	1491	218	1708	5656
PLB1	NM_153021.4	100	1	1385	1467	981	2365	5148
PLCG2	NM_002661.4	99	1	3798	3798	215	4012	8707
PLEKHB1	NM_001130035.1	100	1	570	570	230	799	2027
PNP	NM_000270.3	100	1	867	867	147	1013	2438
POLR3F	NM_006466.3	99	1	951	951	119	1069	2159
PPAPDC2	NM_203453.3	97	17	453	883	10	444	2981
PPARA	XM_006724271.1	100	1	777	777	398	1174	1751
PPARD	NM_177435.2	99	1	1086	1086	310	1395	2028
PPP1R12B	NM_001167858.1	100	1	1161	1161	151	1311	1751
PPP1R7	NM_001282410.1	100	1	843	843	839	1681	1803
PRKACA	NM_002730.3	100	1	1056	1056	201	1256	2689
PRKCA	NM_002737.2	100	1	1032	1032	1032	2063	8787
PRRG2	NM_000951.2	99	15	540	540	249	774	1411
PSG11	NM_001113410.1	99	1	660	660	90	749	1175
PSMC1	NM_002802.2	99	1	1323	1323	49	1371	1586
PSMD3	NM_002809.3	99	1	1605	1605	204	1808	2187
PSMD9	NM_002813.6	100	1	672	672	127	798	2368
PTK2								
ORF #1	NM_153831.3	99	59	2043	2043	1485	3469	4561
ORF #2	NM_153831.3	100	1	2230	3021	311	2540	4561
ORF #2	NM_153831.3	100	2228	3019	3021	2676	3467	4561
PTK2B	NM_173176.2	99	1	3030	3030	241	3270	4180
PTK6	NM_005975.3	100	1	1356	1356	57	1412	2535
PTK7	NM_152880.3	96	41	813	924	1881	2661	4153
PTPN11	NM_080601.1	100	1	1383	1383	381	1763	2069
PTPN6	NM_080548.4	100	1	1794	1794	150	1943	2234
PTTG1	NM_004219.3	99	1	609	609	97	705	786
RAB2B	NM_001163380.1	100	1	36	456	180	215	2883
RAB2B	NM_001163380.1	100	35	456	456	258	679	2883

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
RABGAP1	NM_012197.3	100	1	2992	2994	351	3342	4999
RABGAP1L								
ORF #1	XM_005245681.1	99	1	1599	1818	278	1876	8629
ORF #2	NM_001243763.1	100	1	219	219	225	443	740
ORF #3	NM_005684.4†	100	1	1086	1086	39	1124	1472
RAD54B	NM_001256141.1†	100	1	477	477	508	984	5418
RAI2	NM_021785.4	99	1	1593	1593	265	1857	2229
RALBP1	NM_006788.3	100	1	1968	1968	216	2183	4368
RANBP3	NM_007320.2	100	1	1050	1194	228	1277	3211
RASA1	NM_022650.2	100	1	2613	2613	125	2737	3796
RASSF1	NM_170713.2	100	1	811	813	146	956	1770
RB1	NM_000321.2	100	1	2787	2787	167	2953	4772
RBM4	NM_002896.3	100	1	1095	1095	149	1243	1714
RBP7	NM_052960.2	100	1	405	405	63	467	704
RCAN3	NM_001251979.1	99	1	726	726	173	898	2646
RIN1	NM_004292.2	99	1	2352	2352	128	2479	2698
RIN3								
ORF #1	NM_024832.3	100	1	502	522	153	654	3869
ORF #2	NM_024832.3	100	1	743	885	153	895	3869
ORF #2	NM_024832.3	100	735	772	885	3073	3110	3869
ORF #3	NM_024832.3	99	1	2955	2958	153	3110	3869
RP13-36C9.6	NM_001291529.1†	99	1	570	570	91	660	1166
RPAIN	NM_001160246.1	100	1	321	321	571	891	1362
RPS6	NM_001010.2	100	1	750	750	43	792	829
RTDR1	NM_014433.2	99	1	1047	1047	159	1205	1286
RYBP								
ORF #1	NM_012234.6	99	1	687	687	184	870	4678
ORF #2	NM_012234.6	99	1	687	687	184	870	4678
SCAPER	NM_001145923.1	100	1	879	879	3527	4405	4874
SCLT1	NM_001300897.1	100	1	240	240	505	744	1793

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
SCOC	NM_001153635.1	100	1	366	366	46	411	1873
SDCCAG8	XM_005273023.2	100	1	1083	1083	149	1231	1702
SEMA4D	NM_006378.3	99	1	2589	2589	573	3161	4642
SEPT6								
ORF #1	NM_015129.5	100	1	1305	1305	266	1570	3387
ORF #2	NM_015129.5	99	1	1305	1305	266	1570	3387
SERPINA5	NM_000624.5	99	1	1221	1221	236	1456	2352
SFMBT1	NM_016329.3	99	1	2601	2601	389	2989	4558
SH2B1	NM_001145796.1	99	1	1281	1281	1008	2288	3046
SH2B3	NM_001291424.1	100	1	156	156	2809	2964	4500
SH2D1A	NM_002351.4	100	1	387	387	362	748	2523
SH2D1B								
ORF #1	NM_053282.4	100	1	199	234	123	321	2553
ORF #1	NM_053282.4	100	198	234	234	485	521	2553
ORF #2	NM_053282.4	100	1	397	399	123	519	2553
ORF #3	NM_053282.4	100	1	397	399	123	519	2553
SH2D2A	NM_001161444.1	99	1	1170	1170	141	1310	1445
SH2D3A	NM_005490.2	100	1	1731	1731	195	1925	2376
SH2D4A	NM_001174160.1	99	1	819	819	620	1438	3031
SH3BP2								
ORF #1	NM_001122681.1	99	1	1684	1686	121	1804	9068
ORF #2	XM_006713911.1	99	17	817	818	947	1748	2348
SHD	NM_020209.3	100	1	1023	1023	1464	2486	2598
SHE	NM_001010846.2	99	1	1488	1488	25	1512	6186
SLC39A13								
ORF #1	NM_001128225.2	99	1	1116	1116	177	1292	2429
ORF #2	NM_152264.4	99	1	1095	1095	198	1292	2429
SLPI	NM_003064.3	100	1	399	399	22	420	625
SMARCD1	NM_003076.4	100	1	1548	1548	171	1718	3431
SOCS1	NM_003745.1	100	35	668	697	155	788	1216

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
SOCS3	NM_003955.4	100	1	678	678	419	1096	2737
SOCS4	NM_080867.2	100	1	1323	1323	437	1759	6900
SOCS6	NM_004232.3	100	1	1608	1608	317	1924	5846
SPATA12	NM_181727.1	100	1	573	573	676	1248	2430
SPINK2	NM_021114.3	100	1	255	255	208	462	763
SPRY2	NM_005842.2	99	1	1062	1062	382	1444	2126
SPRY4	NM_001127496.1	100	1	900	900	260	1159	4941
SPSB2	NM_001146316.1	100	1	792	792	157	948	1220
SPSB3	NM_080861.3	97	16	678	849	843	1503	1553
SPTB	NM_001024858.2	97	17	546	805	5984	6513	10074
SRC								
ORF #1	NM_198291.2	100	1	1611	1611	375	1985	4056
ORF #2	NM_198291.2	100	1	1611	1611	375	1985	4056
SRMS	NM_080823.3	99	35	1508	1528	42	1515	2402
SRP19	NM_003135.2	99	1	435	435	190	624	1502
STAT1	NM_139266.2	100	1	2138	2139	389	2526	2798
STAT2	NM_005419.3	100	1	2556	2556	204	2759	4576
STAT3								
ORF #1	NM_003150.3	100	1	2310	2310	219	2528	4953
ORF #2	NM_139276.2	100	1	2313	2313	241	2553	4978
STAT4	NM_001243835.1	100	1	2247	2247	265	2511	2789
STAT5A	NM_003152.3	99	1	2385	2385	643	3027	4314
STK36	NM_015690.4	99	1	3948	3948	280	4227	4946
STRN4	NM_013403.2	100	1	2262	2262	34	2295	3221
SUV39H1	NM_003173.3	100	1	1239	1239	53	1291	2752
SYK								
ORF #1	NM_001174167.2	99	1	1908	1908	202	2109	5075
ORF #2	NM_001174168.2	100	1	1839	1839	121	1959	4925
SYMPK	NM_004819.2	100	1	1599	1602	246	1844	4188
TARBP2	NM_134323.1	100	1	1101	1101	489	1589	1888

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
TCAP	NM_003673.3	99	1	504	504	15	518	963
TDG	NM_003211.4	99	1	1233	1233	224	1456	3251
TEC	NM_003215.2	100	1	1896	1896	92	1987	3620
TENC1	NM_198316.1	100	1	3858	3858	336	4193	4690
TERF2	XM_005256123.1	100	1	714	756	141	854	2977
TGM1	NM_000359.2	100	1	2454	2454	125	2578	2777
TM4SF19	NM_138461.3	99	1	630	630	159	788	1077
TMEM128	NM_032927.3	100	1	426	426	618	1043	1785
TMEM168	NM_022484.5	100	1	1281	1281	1206	2486	7317
TNFAIP1	NM_021137.4	100	1	951	951	450	1400	3822
TNK1	NM_003985.4	99	1	1986	1986	233	2218	2912
TNK2	XM_006713468.1	99	1	1452	1587	796	2247	2511
TNK2	XM_006713468.1	100	1478	1587	1587	2247	2356	2511
TNNC2	NM_003279.2	99	1	483	483	67	549	698
TNS1	NM_022648.4	99	40	1148	1251	345	1453	10276
TNS4	NM_032865.5	99	1	1442	1443	870	2311	4090
TP53	NM_001126112.2	99	1	1180	1182	200	1379	2588
TP53RK	NM_033550.3	100	1	762	762	224	985	3384
TRDN	NM_001256022.1	99	1	504	504	319	822	1413
TRIT1	XM_006710706.1	99	1	975	975	512	1492	2185
TSC1								
ORF #1	NM_001162426.1	100	1	1030	1101	235	1264	8623
ORF #2	XM_006717272.1	100	1	1030	1101	235	1264	2286
TSPAN2	NM_005725.4	99	1	666	666	69	734	3213
TSSK3	NM_052841.3	100	1	807	807	442	1248	1330
TUBA1C	NM_032704.3	99	1	1350	1350	101	1450	1581
TWIST1	NM_000474.3	100	1	609	609	352	960	1669
TWIST2	NM_057179.2	100	1	483	483	185	667	1186
TWSG1	NM_020648.5	100	1	672	672	192	863	3779
TXK	NM_003328.2	100	1	1582	1584	87	1668	2914

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
TYK2	NM_003331.4	99	1	3562	3564	379	3940	4262
UBD	NM_006398.3	99	1	498	498	225	722	1006
USP2	NM_001243759.1	99	1	1089	1089	210	1298	2933
USP46	NM_022832.3	100	1	1101	1101	186	1286	7981
VAC14	NM_018052.3	100	1	2349	2349	259	2607	3107
VASP								
ORF #1	XM_005259199.1	100	1	321	321	331	651	2282
ORF #2	NM_003370.3	100	1	1143	1143	343	1485	2298
VCP								
ORF #1	NM_007126.3	100	1	2420	2421	390	2809	3859
ORF #2	NM_007126.3	100	1	2420	2421	390	2809	3859
VPS39	NM_015289.3	100	1	2154	2154	648	2801	4873
VSTM2L	NM_080607.2	99	1	615	615	255	869	2018
WBSCR27	NM_152559.2	99	1	738	738	41	778	958
WDFY3	NM_014991.4	99	1	846	846	10144	10989	14329
WDR20	NM_144574.3	99	1	1710	1710	73	1782	2446
WDR23	NM_181357.2	100	17	637	662	1099	1719	3663
WDR40B	NM_178470.4	99	1	1392	1392	252	1643	3403
WDR42A								
ORF #1	NM_015726.3	100	1	724	822	213	936	3901
ORF #2	NM_015726.3	99	1	724	822	213	936	3901
WDR83	NM_032332.3	100	1	948	948	89	1036	1219
WIZ	XM_005260007.2	99	15	894	894	1911	2790	6763
WRNIP1	NM_130395.2	99	1	1923	1923	192	2114	2592
WWOX	NM_016373.3	100	1	518	705	367	884	2505
WWOX	NM_016373.3	100	516	704	705	1422	1610	2505
XAGE2XX	NM_130777.2†	99	1	336	336	214	549	651
YES1	NM_005433.3	100	1	1632	1632	222	1853	4685
YPEL3	NM_031477.4	99	1	606	606	586	1192	1588
YWHAQ	NM_006826.3	100	1	737	738	198	934	2272

			ORF			RefSeq		
ORF	**RefSeqRNA**	**Identity**	**St**	**End**	**Len**	**St**	**End**	**Len**
ZAP70								
ORF #1	NM_001079.3	99	1	1860	1860	208	2067	2450
ORF #2	NM_001079.3	100	23	1480	1482	608	2065	2450
ZBTB24	NM_014797.2	100	1	2094	2094	169	2262	5597
ZHX3	XM_005260341.2	99	1	2910	2910	403	3312	9788
ZMAT1	NM_001011657.3	99	1	1710	1710	351	2061	3489
ZNF167	NM_001288591.1	100	1	831	831	426	1256	1742
ZNF281	NM_012482.4	99	1	2688	2688	128	2815	4904
ZNF451	XM_005248994.1	100	1	2753	2799	245	2997	5079
ZNF557	NM_001044387.1	100	1	1126	1126	474	1599	5990
ZNF586	NM_017652.3	97	1	162	300	192	353	2235
ZNF639	NM_016331.1	99	1	1458	1458	446	1903	3013
ZNF655	NM_001009960.1	99	1	1476	1476	221	1696	4585
ZNF670	NM_033213.4	100	1	1170	1170	218	1387	4198
ZNF71	NM_021216.4	99	1	1470	1470	239	1708	3144
ZNF767	XM_005249956.1†	97	7	356	357	308	657	3872
ZNF829	NM_001037232.3	100	1	1299	1299	366	1664	5056
ZNHIT1	NM_006349.2	100	1	465	465	493	957	1208
ZNRD1	NM_001278785.1	99	1	381	381	161	541	760
ZSCAN1	NM_182572.3	99	1	370	537	248	617	2054

Appendix B

List of Protein-Protein Interactions

For each protein-protein interaction detected in this study phosphotyrosine-dependency (pYd) and the collected evidence are indicated. For greater clarity, each (non-self-)interaction is listed twice. Y2H = Yeast two-hybrid; bp = bait-prey; pb = prey-bait; mm = multimer; PCA = protein complementation assay; CoIP = co-immunoprecipitation from mammalian cell culture; pos = positive; neg = negative.

Interaction		pYd	Evidence
ABL2 (KD)	CRK (ORF #1)	Yes	Y2H(pb)
ABL2 (WT)	PIK3R3 (ORF #2)	Yes	Y2H(bp)
ABL2 (KD)	STAT3 (ORF #1)	Yes	Y2H(pb)
ACAP1	GRB2 (ORF #2)	Yes	Y2H(pb)
ACAP1	RABGAP1L (ORF #1)	Yes	Y2H(pb)
ADCK5	TENC1	No	Y2H(pb)
AHDC1	DOK5	No	Y2H(pb)
AKT1	APPL1 (ORF #1)	No	Y2H(pb)
ANAPC10	MAPK8IP2	No	Y2H(pb)
ANGPT1	MAPK8IP2	Yes	Y2H(pb), CoIP(pos)
ANKRD50	MIST	No	Y2H(pb)
ANKS1A	ARHGAP10	No	Y2H(bp)
ANKS1A	DOK3 (ORF #1)	No	Y2H(bp)

Interaction		pYd	Evidence
ANKS1A	KLF15	No	Y2H(bp)
ANKS1A	NHLRC2	No	Y2H(bp)
ANKS1A	PIK3R3 (ORF #5)	Yes	Y2H(bp)
ANKS1A	RANBP3	No	Y2H(bp)
ANKS1A	RB1	No	Y2H(bp)
ANKS1A	SMARCD1	No	Y2H(bp)
ANKS1A	TRIT1	No	Y2H(bp)
ANKS1B	RIN1	No	Y2H(pb)
ANKZF1	CRK (ORF #3)	No	Y2H(pb)
ANLN	CRK (ORF #3)	No	Y2H(pb)
AP1S2	TENC1	No	Y2H(pb)
APBB3	DAPP1 (ORF #1)	No	Y2H(pb), CoIP(pos)
APBB3	PDPK1 (ORF #1)	Yes	Y2H(bp), CoIP(neg)
APPL1 (ORF #1)	AKT1	No	Y2H(bp)
APPL1 (ORF #2)	BATF3	Yes	Y2H(bp)
APPL1 (ORF #2)	BRWD1	Yes	Y2H(bp)
APPL1 (ORF #1)	C22orf28	Yes	Y2H(bp), CoIP(pos)
APPL1 (ORF #2)	CBL	Yes	Y2H(bp)
APPL1 (ORF #2)	CBLB (ORF #2)	Yes	Y2H(bp)
APPL1 (ORF #2)	DOK2 (ORF #2)	Yes	Y2H(bp)
APPL1 (ORF #2)	DOK3 (ORF #2)	Yes	Y2H(bp)
APPL1 (ORF #2)	DOK7	Yes	Y2H(bp)
APPL1 (ORF #2)	ID1	Yes	Y2H(bp)
APPL1 (ORF #2)	INO80E	Yes	Y2H(bp)
APPL1 (ORF #2)	KLF15	Yes	Y2H(bp)
APPL1 (ORF #2)	MAGEC3	Yes	Y2H(bp)
APPL1 (ORF #2)	PIK3R1	Yes	Y2H(bp)
APPL1 (ORF #1)	PIK3R2	No	Y2H(bp), CoIP(neg)
APPL1 (ORF #1)	SCAPER	No	Y2H(bp)
APPL1	SH2D2A	Yes	
ORF #1	SH2D2A		CoIP(neg)

Interaction		pYd	Evidence
ORF #2	SH2D2A		Y2H(bp)
APPL1 (ORF #2)	SOCS6	Yes	Y2H(bp)
APPL1 (ORF #1)	TP53	No	Y2H(bp)
APPL1 (ORF #2)	ZNF829	Yes	Y2H(bp)
APPL2	ATF3	No	Y2H(bp)
APPL2	C17orf82	No	Y2H(bp)
APPL2	CRBN	No	Y2H(bp)
APPL2	CRLF3	No	Y2H(bp)
APPL2	DOK3 (ORF #1)	No	Y2H(bp)
APPL2	GBL	Yes	Y2H(bp)
APPL2	LDHAL6B	No	Y2H(bp)
APPL2	LECT1	No	Y2H(bp)
APPL2	LGALS9C	Yes	Y2H(bp), CoIP(neg)
APPL2	LIX1	Yes	Y2H(bp), CoIP(neg)
APPL2	MAPRE3	Yes	Y2H(bp), CoIP(pos)
APPL2	MLL3	Yes	Y2H(bp)
APPL2	PINK1	No	Y2H(bp)
APPL2	RAI2	No	Y2H(bp)
APPL2	RBP7	No	Y2H(bp)
APPL2	TSC1 (ORF #1)	No	Y2H(bp)
ARAP3	MIST	Yes	Y2H(pb)
ARFGAP1	RABGAP1L (ORF #1)	Yes	Y2H(pb), CoIP(neg)
ARHGAP10	ANKS1A	No	Y2H(pb)
ARHGAP10	SH3BP2 (ORF #1)	Yes	Y2H(pb)
ARID5A	SH2D2A	No	Y2H(pb)
ARL6IP4	GRB2 (ORF #2)	No	Y2H(pb), PCA(pos), CoIP(pos)
ARL6IP4	PIK3R3	Yes	
ARL6IP4	ORF #1		PCA(pos)
ARL6IP4	ORF #2		Y2H(pb), CoIP(pos)
ASB3	RIN3	No	
ASB3	ORF #1		Y2H(pb)

Interaction		pYd	Evidence
ASB3	ORF #2		Y2H(pb)
ASB3	SH2D2A	Yes	Y2H(pb), CoIP(pos)
ASB3	TENC1	No	Y2H(pb)
ASB8	RIN3 (ORF #1)	No	Y2H(pb)
ASB9	CRK (ORF #3)	No	Y2H(pb), CoIP(pos)
ASB9	PLCG2	No	Y2H(pb), CoIP(neg)
ASCL4	CRK (ORF #2)	No	Y2H(pb)
ATF3	APPL2	No	Y2H(pb)
ATF3	CRK (ORF #2)	No	Y2H(pb)
ATF3	DOK5	No	Y2H(pb)
ATF3	SH2D1A	Yes	Y2H(pb)
ATF3	STAT3 (ORF #2)	No	Y2H(pb)
BAIAP2	EPS8	No	
ORF #1	EPS8		Y2H(pb)
ORF #2	EPS8		Y2H(pb)
BAIAP2 (ORF #2)	GRB10	Yes	Y2H(pb)
BAT2L	CRK (ORF #3)	No	Y2H(pb)
BATF3	APPL1 (ORF #2)	Yes	Y2H(pb)
BATF3	CRK	No	
BATF3	ORF #1		Y2H(pb)
BATF3	ORF #2		Y2H(pb)
BATF3	STAT3 (ORF #2)	No	Y2H(pb)
BCAR3	IRS1	Yes	Y2H(pb), CoIP(pos)
BECN1	BTK (ORF #2)	No	Y2H(pb)
BECN1	FRS2	No	Y2H(pb)
BECN1	SHD	No	Y2H(pb)
BECN1	YES1	Yes	Y2H(pb), CoIP(neg)
BEX5	CRK (ORF #2)	Yes	Y2H(pb)
BMX	MIST	No	
ORF #2	MIST		Y2H(pb)
ORF #3	MIST		Y2H(pb)

Interaction		pYd	Evidence
BMX (ORF #2)	STAT3 (ORF #2)	No	Y2H(pb)
BRWD1	APPL1 (ORF #2)	Yes	Y2H(pb)
BRWD1	STAT3 (ORF #1)	Yes	Y2H(pb)
BTK (ORF #2)	BECN1	No	Y2H(bp)
BTK (ORF #2)	MRPS22	No	Y2H(bp)
BTK (ORF #2)	PIK3R3 (ORF #2)	Yes	Y2H(pb)
C10orf18	STAT3	Yes	
C10orf18	ORF #1		Y2H(pb)
C10orf18	ORF #2		Y2H(pb)
C10orf81	GRB2 (ORF #2)	Yes	Y2H(pb), PCA(pos), CoIP(pos)
C10orf81	PIK3R3	Yes	
C10orf81	ORF #1		PCA(pos)
C10orf81	ORF #2		Y2H(pb), CoIP(pos)
C17orf53	PIK3R3 (ORF #2)	No	Y2H(pb)
C17orf82	APPL2	No	Y2H(pb)
C17orf82	SH2D2A	No	Y2H(pb)
C19orf66	TENC1	No	Y2H(pb)
C1orf135	GRB2 (ORF #2)	Yes	Y2H(pb)
C1orf135	PIK3R3	No	
C1orf135	ORF #1		Y2H(pb)
C1orf135	ORF #2		CoIP(neg)
C1orf62	NOS1AP	No	Y2H(pb)
C1QTNF3	MAPK8IP2	No	Y2H(pb)
C21orf77	TENC1	No	Y2H(pb)
C22orf28	APPL1 (ORF #1)	Yes	Y2H(pb), CoIP(pos)
C22orf28	CRK (ORF #3)	Yes	Y2H(pb), CoIP(neg)
C22orf29	PLCG2	No	Y2H(pb)
C22orf39	MAPK8IP2	Yes	Y2H(pb), CoIP(neg)
C3orf10	PIK3R3 (ORF #2)	No	Y2H(pb)
C3orf34	PIK3R3 (ORF #2)	No	Y2H(pb)
C4orf17	CRK (ORF #2)	No	Y2H(pb)

Interaction		pYd	Evidence
C5orf35	CRK (ORF #3)	No	Y2H(pb)
C5orf35	PLCG2	No	Y2H(pb)
C6orf125	GRB2 (ORF #2)	Yes	Y2H(pb)
C6orf125	RABGAP1L (ORF #1)	Yes	Y2H(pb)
C6orf141	CRK (ORF #2)	No	Y2H(pb)
C6orf141	EPS8	No	Y2H(pb)
C6orf146	MIST	Yes	Y2H(pb)
C8orf33	MAPK8IP2	Yes	Y2H(pb), CoIP(neg)
C9orf43	RABGAP1L (ORF #1)	Yes	Y2H(pb), CoIP(pos)
CA8	STAT3 (ORF #2)	Yes	Y2H(pb)
CAMK1	PIK3R3 (ORF #2)	No	Y2H(pb)
CAPN10	SHD	Yes	Y2H(pb)
CBL	APPL1 (ORF #2)	Yes	Y2H(pb)
CBL	CBL	No	Y2H(mm)
CBL	CRK	Yes	
CBL	ORF #1		Y2H(pb)
CBL	ORF #2		Y2H(pb)
CBL	CRKL (ORF #2)	No	Y2H(bp)
CBL	EPS8	Yes	Y2H(pb)
CBL	FYN (ORF #2)	No	Y2H(bp)
CBL	GRAP2 (ORF #1)	No	Y2H(pb)
CBL	GRB2 (ORF #2)	No	Y2H(pb), CoIP(pos)
CBL	MZF1	No	Y2H(bp)
CBL	PIK3R1	Yes	Y2H(bp)
CBL	PIK3R2	Yes	Y2H(pb), CoIP(neg)
CBL	PIK3R3	Yes	
CBL	ORF #1		Y2H(bp)
CBL	ORF #2		Y2H(bp & pb), CoIP(pos)
CBL	PRKCA	No	Y2H(bp)
CBL	SERPINA5	No	Y2H(bp)
CBL	STAT3 (ORF #2)	Yes	Y2H(pb)

Interaction		pYd	Evidence
CBL	STAT5A	Yes	Y2H(pb), CoIP(neg)
CBLB (ORF #2)	APPL1 (ORF #2)	Yes	Y2H(pb)
CBLB	CRK	Yes	
ORF #1	ORF #1		Y2H(bp)
ORF #1	ORF #3		CoIP(neg)
ORF #2	ORF #1		Y2H(pb)
ORF #2	ORF #2		Y2H(pb)
CBLB (ORF #1)	CRKL (ORF #2)	Yes	Y2H(bp)
CBLB (ORF #1)	FYN (ORF #2)	No	Y2H(pb), CoIP(neg)
CBLB (ORF #2)	GRAP2 (ORF #1)	No	Y2H(pb)
CBLC	CRK (ORF #2)	Yes	Y2H(pb)
CCDC14	PIK3R3 (ORF #2)	Yes	Y2H(pb)
CCDC33	PIK3R3 (ORF #2)	No	Y2H(pb)
CCDC74A	SH2D1A	No	Y2H(pb)
CCDC87	STAT3 (ORF #2)	Yes	Y2H(pb)
CCM2	DOK4	Yes	Y2H(pb), CoIP(pos)
CCM2	KRIT1	No	Y2H(bp)
CCM2	TWIST2	Yes	Y2H(bp), CoIP(pos)
CCM2	VSTM2L	Yes	Y2H(bp), CoIP(pos)
CDCA4	DOK5	No	Y2H(pb)
CENPB	SH2D3A	No	Y2H(pb)
CHN2	RANBP3	No	Y2H(bp)
CHN2	RB1	No	Y2H(bp)
CHN2	SMARCD1	No	Y2H(bp)
CHN2	TP53RK	No	Y2H(bp)
CHTF18	CRK (ORF #2)	Yes	Y2H(pb)
CHTF18	STAT3	No	
CHTF18	ORF #1		Y2H(pb)
CHTF18	ORF #2		Y2H(pb)
CKM	NOS1AP	No	Y2H(pb)
CNDP2	CRK (ORF #2)	Yes	Y2H(pb)

Interaction		pYd	Evidence
CNDP2	DAPP1 (ORF #2)	No	Y2H(pb)
CNDP2	STAT3	No	
CNDP2	ORF #1		Y2H(pb)
CNDP2	ORF #2		Y2H(pb)
COQ9	MIST	Yes	Y2H(pb)
CORO6	CRK (ORF #3)	No	Y2H(pb)
COX6B2	TENC1	No	Y2H(pb)
CRBN	APPL2	No	Y2H(pb)
CRBN	GRB2 (ORF #2)	Yes	Y2H(pb)
CRBN	PIK3R3 (ORF #2)	No	Y2H(pb)
CRK (ORF #1)	ABL2 (KD)	Yes	Y2H(bp)
CRK (ORF #3)	ANKZF1	No	Y2H(bp)
CRK (ORF #3)	ANLN	No	Y2H(bp)
CRK (ORF #3)	ASB9	No	Y2H(bp), CoIP(pos)
CRK (ORF #2)	ASCL4	No	Y2H(bp)
CRK (ORF #2)	ATF3	No	Y2H(bp)
CRK (ORF #3)	BAT2L	No	Y2H(bp)
CRK	BATF3	No	
ORF #1	BATF3		Y2H(bp)
ORF #2	BATF3		Y2H(bp)
CRK (ORF #2)	BEX5	Yes	Y2H(bp)
CRK (ORF #3)	C22orf28	Yes	Y2H(bp), CoIP(neg)
CRK (ORF #2)	C4orf17	No	Y2H(bp)
CRK (ORF #3)	C5orf35	No	Y2H(bp)
CRK (ORF #2)	C6orf141	No	Y2H(bp)
CRK	CBL	Yes	
ORF #1	CBL		Y2H(bp)
ORF #2	CBL		Y2H(bp)
CRK	CBLB	Yes	
ORF #1	ORF #1		Y2H(pb)
ORF #1	ORF #2		Y2H(bp)

Interaction		pYd	Evidence
ORF #2	ORF #2		Y2H(bp)
ORF #3	ORF #1		CoIP(neg)
CRK (ORF #2)	CBLC	Yes	Y2H(bp)
CRK (ORF #2)	CHTF18	Yes	Y2H(bp)
CRK (ORF #2)	CNDP2	Yes	Y2H(bp)
CRK (ORF #3)	CORO6	No	Y2H(bp)
CRK	DOK2	Yes	
ORF #1	ORF #2		Y2H(bp)
ORF #2	ORF #2		Y2H(bp)
CRK	DOK3	Yes	
ORF #1	ORF #2		Y2H(bp)
ORF #2	ORF #1		Y2H(bp)
ORF #2	ORF #2		Y2H(bp)
CRK (ORF #2)	DOK7	Yes	Y2H(bp)
CRK (ORF #3)	ELK1	Yes	Y2H(bp), CoIP(neg)
CRK (ORF #3)	EPYC (ORF #2)	Yes	Y2H(bp)
CRK (ORF #3)	ESD	No	Y2H(bp)
CRK (ORF #2)	EYA3	Yes	Y2H(bp)
CRK	FAM59A	Yes	
ORF #1	FAM59A		Y2H(bp)
ORF #2	FAM59A		Y2H(bp)
CRK (ORF #1)	FER	No	Y2H(pb)
CRK (ORF #3)	FSTL1	Yes	Y2H(bp)
CRK (ORF #2)	GABPB2 (ORF #2)	Yes	Y2H(bp)
CRK (ORF #3)	HABP4	No	Y2H(bp)
CRK (ORF #2)	HSH2D (ORF #1)	Yes	Y2H(bp)
CRK (ORF #3)	IFT140	No	Y2H(bp)
CRK (ORF #2)	INO80E	No	Y2H(bp)
CRK (ORF #3)	ISL1	No	Y2H(bp)
CRK	KCTD13	Yes	
ORF #1	KCTD13		Y2H(bp)

Interaction		pYd	Evidence
ORF #2	KCTD13		Y2H(bp)
CRK (ORF #3)	KCTD17	No	Y2H(bp)
CRK (ORF #2)	KLF15	No	Y2H(bp)
CRK (ORF #2)	KLHL20	Yes	Y2H(bp)
CRK	LASP1	Yes	
ORF #1	LASP1		Y2H(bp)
ORF #2	LASP1		Y2H(bp)
ORF #3	LASP1		CoIP(neg)
CRK (ORF #3)	LHX8	No	Y2H(bp)
CRK (ORF #2)	MAGEC3	Yes	Y2H(bp)
CRK (ORF #1)	MIST	No	Y2H(pb)
CRK	MNDA	No	
ORF #1	MNDA		Y2H(bp)
ORF #2	MNDA		Y2H(bp)
CRK (ORF #3)	MPG	Yes	Y2H(bp)
CRK (ORF #2)	MYLIP	Yes	Y2H(bp)
CRK (ORF #2)	MYOZ2	Yes	Y2H(bp)
CRK (ORF #2)	NUFIP2	Yes	Y2H(bp)
CRK (ORF #2)	OFCC1	Yes	Y2H(bp)
CRK (ORF #2)	PAFAH1B2	Yes	Y2H(bp)
CRK (ORF #2)	PHC2	Yes	Y2H(bp)
CRK	PIK3R1	Yes	
ORF #1	PIK3R1		Y2H(bp)
ORF #2	PIK3R1		Y2H(bp)
ORF #3	PIK3R1		CoIP(neg)
CRK	PIK3R2	Yes	
ORF #2	PIK3R2		Y2H(bp)
ORF #3	PIK3R2		CoIP(neg)
CRK	PIK3R3	Yes	
ORF #1	ORF #3		Y2H(bp)
ORF #1	ORF #5		Y2H(bp)

Interaction		pYd	Evidence
ORF #2	ORF #3		Y2H(bp)
ORF #2	ORF #4		Y2H(bp)
ORF #2	ORF #5		Y2H(bp)
CRK (ORF #2)	PRKACA	Yes	Y2H(bp)
CRK (ORF #3)	PRRG2	Yes	Y2H(bp)
CRK (ORF #2)	PSMC1	No	Y2H(bp)
CRK (ORF #1)	PTK2 (ORF #1 KD)	Yes	Y2H(bp)
CRK (ORF #3)	PTTG1	No	Y2H(bp)
CRK (ORF #2)	RAB2B	Yes	Y2H(bp)
CRK (ORF #4)	RIN3 (ORF #3)	No	Y2H(pb)
CRK	RYBP	No	
ORF #1	ORF #2		Y2H(bp)
ORF #2	ORF #2		Y2H(bp)
CRK (ORF #3)	SEMA4D	Yes	Y2H(bp), CoIP(neg)
CRK	SEPT6	No	
ORF #1	ORF #1		Y2H(bp)
ORF #1	ORF #2		Y2H(bp)
ORF #2	ORF #2		Y2H(bp)
CRK	SH2D2A	Yes	
ORF #1	SH2D2A		Y2H(bp)
ORF #2	SH2D2A		Y2H(bp)
ORF #3	SH2D2A		CoIP(neg)
CRK (ORF #2)	SOCS1	Yes	Y2H(bp)
CRK (ORF #2)	SOCS6	Yes	Y2H(bp)
CRK (ORF #3)	STAT4	Yes	Y2H(bp), CoIP(neg)
CRK (ORF #3)	STRN4	No	Y2H(bp)
CRK (ORF #2)	TCAP	Yes	Y2H(bp)
CRK (ORF #2)	TDG	No	Y2H(bp)
CRK (ORF #2)	TM4SF19	Yes	Y2H(bp)
CRK (ORF #3)	TP53	No	Y2H(bp)
CRK (ORF #3)	TUBA1C	No	Y2H(bp)

Interaction		pYd	Evidence
CRK (ORF #3)	TWIST2	No	Y2H(bp), CoIP(neg)
CRK (ORF #1)	VAC14	Yes	Y2H(bp)
CRK	WDR83	No	
ORF #1	WDR83		Y2H(bp)
ORF #2	WDR83		Y2H(bp)
CRK (ORF #3)	ZNF167	No	Y2H(bp)
CRK (ORF #2)	ZNF557	No	Y2H(bp)
CRKL (ORF #2)	CBL	No	Y2H(pb)
CRKL (ORF #2)	CBLB (ORF #1)	Yes	Y2H(pb)
CRKL (ORF #1)	DOK1	Yes	Y2H(pb)
CRKL (ORF #1)	DOK2 (ORF #1)	Yes	Y2H(pb)
CRKL (ORF #1)	KIAA0317	Yes	Y2H(bp)
CRKL (ORF #1)	RIN3 (ORF #3)	No	Y2H(pb)
CRKL (ORF #1)	SOCS1	Yes	Y2H(bp)
CRKL (ORF #1)	TENC1	No	Y2H(pb)
CRKL (ORF #1)	TMEM168	Yes	Y2H(bp)
CRLF3	APPL2	No	Y2H(pb)
CRP	MAPK8IP2	No	Y2H(pb)
CRYBA2	PIK3R3 (ORF #2)	Yes	Y2H(pb), CoIP(pos)
CSAD	SH2B1	Yes	Y2H(pb)
CSK	PDPK1 (ORF #2)	Yes	Y2H(bp), CoIP(neg)
CSRP3	RIN3	No	
CSRP3	ORF #1		Y2H(pb)
CSRP3	ORF #2		Y2H(pb)
CTNNBL1	MAPK8IP2	No	Y2H(pb)
CYB561D2	MIST	Yes	Y2H(pb)
CYP2C8	RIN3 (ORF #1)	No	Y2H(pb)
CYP46A1	RIN3 (ORF #1)	No	Y2H(pb)
DAPP1 (ORF #1)	APBB3	No	Y2H(bp), CoIP(pos)
DAPP1 (ORF #2)	CNDP2	No	Y2H(bp)
DAPP1 (ORF #2)	GSTCD	Yes	Y2H(bp)

Interaction		pYd	Evidence
DAPP1 (ORF #2)	MYLIP	Yes	Y2H(bp)
DAPP1 (ORF #2)	OSAP	No	Y2H(bp)
DAPP1 (ORF #1)	PLAGL2	No	Y2H(bp), CoIP(neg)
DAPP1 (ORF #1)	RBP7	No	Y2H(bp)
DAPP1 (ORF #2)	TM4SF19	No	Y2H(bp)
DAPP1 (ORF #2)	TNFAIP1	Yes	Y2H(bp)
DAPP1 (ORF #1)	WDR20	Yes	Y2H(bp), CoIP(pos)
DAPP1 (ORF #1)	ZNHIT1	Yes	Y2H(bp)
DBN1	GRB2 (ORF #2)	Yes	Y2H(pb), CoIP(pos)
DBN1	RABGAP1L (ORF #1)	Yes	Y2H(pb), CoIP(neg)
DDIT4L	TENC1	No	Y2H(pb)
DDX5	FRS3 (ORF #2)	No	Y2H(pb)
DOK1	CRKL (ORF #1)	Yes	Y2H(bp)
DOK1	DOK3 (ORF #1)	Yes	Y2H(bp)
DOK2 (ORF #2)	APPL1 (ORF #2)	Yes	Y2H(pb)
DOK2	CRK	Yes	
ORF #2	ORF #1		Y2H(pb)
ORF #2	ORF #2		Y2H(pb)
DOK2 (ORF #1)	CRKL (ORF #1)	Yes	Y2H(bp)
DOK2 (ORF #1)	DOK3 (ORF #1)	Yes	Y2H(bp)
DOK2 (ORF #1)	LCK (ORF #2)	Yes	Y2H(bp)
DOK2 (ORF #2)	STAT3 (ORF #1)	Yes	Y2H(pb)
DOK3 (ORF #1)	ANKS1A	No	Y2H(pb)
DOK3 (ORF #2)	APPL1 (ORF #2)	Yes	Y2H(pb)
DOK3 (ORF #1)	APPL2	No	Y2H(pb)
DOK3	CRK	Yes	
ORF #1	ORF #2		Y2H(pb)
ORF #2	ORF #1		Y2H(pb)
ORF #2	ORF #2		Y2H(pb)
DOK3 (ORF #1)	DOK1	Yes	Y2H(pb)
DOK3 (ORF #1)	DOK2 (ORF #1)	Yes	Y2H(pb)

Interaction		pYd	Evidence
DOK3 (ORF #1)	DOK4	Yes	Y2H(pb), CoIP(pos)
DOK3 (ORF #1)	DOK5	Yes	Y2H(pb)
DOK3 (ORF #1)	RIN3 (ORF #2)	No	Y2H(pb)
DOK3 (ORF #1)	SHE	Yes	Y2H(pb)
DOK3	STAT3	Yes	
ORF #1	ORF #1		Y2H(pb)
ORF #2	ORF #1		Y2H(pb)
ORF #2	ORF #2		Y2H(pb)
DOK3 (ORF #1)	TENC1	No	Y2H(pb)
DOK4	CCM2	Yes	Y2H(bp), CoIP(pos)
DOK4	DOK3 (ORF #1)	Yes	Y2H(bp), CoIP(pos)
DOK4	LIX1	No	Y2H(bp)
DOK4	MAPRE3	No	Y2H(bp)
DOK4	OLIG1	No	Y2H(bp), CoIP(neg)
DOK4	PTK2 (ORF #1 WT)	Yes	Y2H(bp), CoIP(neg)
DOK4	RBP7	No	Y2H(bp)
DOK4	STAT2	Yes	Y2H(bp), CoIP(neg)
DOK5	AHDC1	No	Y2H(bp)
DOK5	ATF3	No	Y2H(bp)
DOK5	CDCA4	No	Y2H(bp)
DOK5	DOK3 (ORF #1)	Yes	Y2H(bp)
DOK5	FANCG	No	Y2H(bp)
DOK5	OLIG1	No	Y2H(bp)
DOK5	RAI2	No	Y2H(bp)
DOK5	SCOC	Yes	Y2H(bp), CoIP(neg)
DOK5	TSC1 (ORF #1)	No	Y2H(bp)
DOK5	WIZ	No	Y2H(bp)
DOK7	APPL1 (ORF #2)	Yes	Y2H(pb)
DOK7	CRK (ORF #2)	Yes	Y2H(pb)
DOK7	EFHC1	No	Y2H(bp)
DOK7	MPZL1	Yes	Y2H(bp)

Interaction		pYd	Evidence
DOK7	YPEL3	No	Y2H(bp)
DTNA	STAT3	Yes	
DTNA	ORF #1		Y2H(pb)
DTNA	ORF #2		Y2H(pb)
DUSP18	MAPK8IP2	No	Y2H(pb)
E2F2	GRB2 (ORF #2)	Yes	Y2H(pb)
E2F6	PIK3R3 (ORF #2)	Yes	Y2H(pb)
EFHC1	DOK7	No	Y2H(pb)
ELK1	CRK (ORF #3)	Yes	Y2H(pb), CoIP(neg)
ELK1	PLCG2	Yes	Y2H(pb), CoIP(neg)
EPB41L4A	TENC1	No	Y2H(pb)
EPS8	BAIAP2	No	
EPS8	ORF #1		Y2H(bp)
EPS8	ORF #2		Y2H(bp)
EPS8	C6orf141	No	Y2H(bp)
EPS8	CBL	Yes	Y2H(bp)
EPS8	RCAN3	No	Y2H(bp), CoIP(neg)
EPS8	RTDR1	No	Y2H(bp)
EPS8	SERPINA5	No	Y2H(bp)
EPS8	XAGE2XX	Yes	Y2H(bp)
EPS8L2	RIN3	No	
EPS8L2	ORF #1		Y2H(pb)
EPS8L2	ORF #2		Y2H(pb)
EPS8L2	TENC1	No	Y2H(pb)
EPYC (ORF #2)	CRK (ORF #3)	Yes	Y2H(pb)
EPYC (ORF #1)	PLCG2	Yes	Y2H(pb)
ESD	CRK (ORF #3)	No	Y2H(pb)
ESD	PLCG2	No	Y2H(pb)
EVI1	PIK3R3 (ORF #2)	Yes	Y2H(pb), CoIP(neg)
EVI1	RABGAP1L	Yes	
EVI1	ORF #1		Y2H(pb)

Interaction		pYd	Evidence
EVI1	ORF #3		CoIP(pos)
EYA3	CRK (ORF #2)	Yes	Y2H(pb)
EZH2	RIN3 (ORF #2)	No	Y2H(pb)
FABP7	TENC1	No	Y2H(pb)
FAM114A1	TENC1	No	Y2H(pb)
FAM117B	STAT3 (ORF #2)	Yes	Y2H(pb)
FAM127C	RIN3 (ORF #1)	No	Y2H(pb)
FAM46A	RIN3	No	
FAM46A	ORF #1		Y2H(pb)
FAM46A	ORF #2		Y2H(pb)
FAM46A	SH2D2A	Yes	Y2H(pb), CoIP(pos)
FAM46A	SYK (ORF #1 WT)	Yes	Y2H(pb), CoIP(pos)
FAM46A	TENC1	No	Y2H(pb)
FAM46B	SH2D2A	Yes	Y2H(pb), CoIP(pos)
FAM59A	CRK	Yes	
FAM59A	ORF #1		Y2H(pb)
FAM59A	ORF #2		Y2H(pb)
FAM59A	GRAP2 (ORF #1)	No	Y2H(pb)
FANCG	DOK5	No	Y2H(pb)
FER	CRK (ORF #1)	No	Y2H(bp)
FER	KCTD4	Yes	Y2H(bp)
FER	PIK3R1	No	Y2H(bp)
FER	PIK3R3 (ORF #5)	No	Y2H(bp)
FER	PPP1R7	No	Y2H(bp)
FGF12	MAPK8IP2	No	Y2H(pb)
FGF13	MAPK8IP2	No	Y2H(pb)
FGF21	TENC1	No	Y2H(pb)
FOXO1	PIK3R3 (ORF #2)	No	Y2H(pb)
FOXO1	TENC1	No	Y2H(pb)
FRS2	BECN1	No	Y2H(bp)
FRS2	RPS6	No	Y2H(bp)

Interaction		pYd	Evidence
FRS3 (ORF #2)	DDX5	No	Y2H(bp)
FRS3	FRS3	No	
ORF #1	ORF #2		Y2H(mm)
ORF #2	ORF #2		CoIP(pos)
FRS3 (ORF #2)	GRB14	No	Y2H(bp)
FRS3	HCK	No	
ORF #1	HCK		Y2H(pb)
ORF #2	HCK		CoIP(pos)
FRS3 (ORF #2)	MATK (ORF #2)	Yes	Y2H(bp), CoIP(neg)
FRS3 (ORF #2)	NUMB (ORF #2)	Yes	Y2H(bp), CoIP(neg)
FRS3 (ORF #2)	PIK3R3 (ORF #2)	Yes	Y2H(bp), CoIP(pos)
FRS3 (ORF #2)	SH2B1	No	Y2H(bp)
FRS3 (ORF #2)	SOCS4	No	Y2H(bp), CoIP(pos)
FRS3 (ORF #2)	SOCS6	No	Y2H(bp)
FRS3 (ORF #2)	TNS1	No	Y2H(bp)
FRS3	VASP	Yes	
ORF #2	ORF #1		Y2H(bp)
ORF #2	ORF #2		CoIP(neg)
FSTL1	CRK (ORF #3)	Yes	Y2H(pb)
FYB	FYN	Yes	
FYB	ORF #2		Y2H(bp)
FYB	SH2		Y2H(bp)
FYB	LCP2	Yes	Y2H(bp)
FYB	NCK2	Yes	Y2H(bp)
FYN (ORF #2)	CBL	No	Y2H(pb)
FYN (ORF #2)	CBLB (ORF #1)	No	Y2H(bp), CoIP(neg)
FYN	FYB	Yes	
ORF #2	FYB		Y2H(pb)
SH2	FYB		Y2H(pb)
FYN (SH2)	NOS1AP	No	Y2H(pb)
FYN (ORF #1 WT)	PIK3R3 (ORF #2)	No	Y2H(pb)

Interaction		pYd	Evidence
GABPB2 (ORF #2)	CRK (ORF #2)	Yes	Y2H(pb)
GABPB2 (ORF #1)	PLCG2	No	Y2H(pb)
GALNT3	MAPK8IP2	No	Y2H(pb)
GATA1	RIN3 (ORF #2)	No	Y2H(pb)
GATA1	TENC1	No	Y2H(pb)
GBA3	TENC1	No	Y2H(pb)
GBL	APPL2	Yes	Y2H(pb)
GLIS3	MIST	Yes	Y2H(pb)
GRAP2 (ORF #1)	CBL	No	Y2H(bp)
GRAP2 (ORF #1)	CBLB (ORF #2)	No	Y2H(bp)
GRAP2 (ORF #1)	FAM59A	No	Y2H(bp)
GRAP2 (ORF #2)	RIN3 (ORF #3)	No	Y2H(pb)
GRB2 (ORF #2)	ACAP1	Yes	Y2H(bp)
GRB2 (ORF #2)	ARL6IP4	No	Y2H(bp), PCA(pos), CoIP(pos)
GRB2 (ORF #2)	C10orf81	Yes	Y2H(bp), PCA(pos), CoIP(pos)
GRB2 (ORF #2)	C1orf135	Yes	Y2H(bp)
GRB2 (ORF #2)	C6orf125	Yes	Y2H(bp)
GRB2 (ORF #2)	CBL	No	Y2H(bp), CoIP(pos)
GRB2 (ORF #2)	CRBN	Yes	Y2H(bp)
GRB2 (ORF #2)	DBN1	Yes	Y2H(bp), CoIP(pos)
GRB2 (ORF #2)	E2F2	Yes	Y2H(bp)
GRB2 (ORF #2)	LMX1A	Yes	Y2H(bp), CoIP(pos)
GRB2 (ORF #1)	MIST	No	Y2H(pb)
GRB2 (ORF #2)	MYOZ1	Yes	Y2H(bp), CoIP(neg)
GRB2 (ORF #2)	OLIG1	No	Y2H(bp), PCA(neg), CoIP(neg)
GRB2 (ORF #2)	PACRGL	Yes	Y2H(bp), CoIP(neg)
GRB2 (ORF #2)	PARD6A	Yes	Y2H(bp), PCA(pos), CoIP(pos)
GRB2 (ORF #2)	PCDHB5	Yes	Y2H(bp)
GRB2 (ORF #2)	PPARA	Yes	Y2H(bp), CoIP(pos)

Interaction		pYd	Evidence
GRB2 (ORF #2)	RB1	Yes	Y2H(bp), PCA(pos), CoIP(pos)
GRB2 (ORF #2)	SOCS4	Yes	Y2H(bp), PCA(neg), CoIP(pos)
GRB2 (ORF #2)	TMEM128	Yes	Y2H(bp), PCA(neg), CoIP(pos)
GRB2 (ORF #2)	TSPAN2	Yes	Y2H(bp), PCA(pos), CoIP(pos)
GRB2	VCP	Yes	
ORF #2	ORF #1		Y2H(bp)
ORF #2	ORF #2		CoIP(pos)
GRB2 (ORF #2)	WBSCR27	No	Y2H(bp), PCA(neg), CoIP(neg)
GRB2 (ORF #2)	WDFY3	Yes	Y2H(bp), CoIP(neg)
GRB2 (ORF #2)	ZMAT1	No	Y2H(bp)
GRB7	OLIG1	Yes	Y2H(bp), CoIP(pos)
GRB7	PTK2 (ORF #1 WT)	Yes	Y2H(bp), CoIP(pos)
GRB7	ZNHIT1	Yes	Y2H(bp)
GRB10	BAIAP2 (ORF #2)	Yes	Y2H(bp)
GRB10	RCAN3	Yes	Y2H(bp), CoIP(pos)
GRB14	FRS3 (ORF #2)	No	Y2H(pb)
GRB14	HCK	No	Y2H(pb)
GRB14	MST4	Yes	Y2H(bp)
GSTCD	DAPP1 (ORF #2)	Yes	Y2H(pb)
GSTCD	MIST	No	Y2H(pb)
GSTCD	STAT3	Yes	
GSTCD	ORF #1		Y2H(pb)
GSTCD	ORF #2		Y2H(pb)
GSTCD	TNS4	Yes	Y2H(pb)
HABP4	CRK (ORF #3)	No	Y2H(pb)
HCK	FRS3	No	
HCK	ORF #1		Y2H(bp)
HCK	ORF #2		CoIP(pos)

Interaction		pYd	Evidence
HCK	GRB14	No	Y2H(bp)
HCK	LETMD1	No	Y2H(bp)
HCK	NGEF	Yes	Y2H(bp)
HCK	OLIG1	No	Y2H(bp), CoIP(pos)
HCK	PIK3R3	Yes	
HCK	ORF #1		Y2H(bp)
HCK	ORF #2		Y2H(bp), CoIP(pos)
HCK	ORF #3		Y2H(bp)
HCK	PTK2 (ORF #1 WT)	No	Y2H(bp), CoIP(pos)
HESX1	STAT3	Yes	
HESX1	ORF #1		Y2H(pb)
HESX1	ORF #2		Y2H(pb)
HLA-C	MAPK8IP2	No	Y2H(pb)
HSH2D (ORF #1)	CRK (ORF #2)	Yes	Y2H(pb)
HSH2D (ORF #2)	LDHAL6B	Yes	Y2H(bp)
HSH2D (ORF #2)	LECT1	Yes	Y2H(bp)
HSH2D (ORF #2)	OLIG1	No	Y2H(bp)
HSH2D	PIK3R3	Yes	
ORF #2	ORF #1		Y2H(bp)
ORF #2	ORF #2		Y2H(bp), CoIP(neg)
ORF #2	ORF #3		Y2H(bp)
HSH2D (ORF #2)	PINK1	Yes	Y2H(bp)
HSH2D (ORF #2)	PPAPDC2	Yes	Y2H(bp)
HSH2D (ORF #2)	RBP7	Yes	Y2H(bp)
HSH2D (ORF #2)	TNK2	No	Y2H(bp)
HSH2D (ORF #2)	TSC1 (ORF #1)	Yes	Y2H(bp)
HSPA1A	MAPK8IP2	No	
ORF #1	MAPK8IP2		Y2H(pb)
ORF #2	MAPK8IP2		CoIP(pos)
HSPA1A	SH2D1B	Yes	
ORF #1	ORF #3		Y2H(pb)

Interaction		pYd	Evidence
ORF #2	ORF #3		CoIP(pos)
HSPD1 (ORF #1)	NOS1AP	No	Y2H(pb)
HSPD1 (ORF #2)	PLCG2	Yes	Y2H(pb)
HUWE1	TENC1	No	Y2H(pb)
ID1	APPL1 (ORF #2)	Yes	Y2H(pb)
ID2	RIN3	No	
ID2	ORF #1		Y2H(pb)
ID2	ORF #2		Y2H(pb)
ID2	TENC1	No	Y2H(pb)
IFT140	CRK (ORF #3)	No	Y2H(pb)
ING4	MAPK8IP2	No	Y2H(pb)
INO80E	APPL1 (ORF #2)	Yes	Y2H(pb)
INO80E	CRK (ORF #2)	No	Y2H(pb)
IQUB	PIK3R3 (ORF #2)	Yes	Y2H(pb)
IRS1	BCAR3	Yes	Y2H(bp), CoIP(pos)
IRS1	NUMB (ORF #2)	Yes	Y2H(bp), CoIP(neg)
IRS1	PELI1	No	Y2H(bp)
IRS1	PIK3R1	Yes	Y2H(bp), CoIP(pos)
IRS1	PIK3R3	Yes	
IRS1	ORF #1		Y2H(bp)
IRS1	ORF #2		Y2H(bp), CoIP(pos)
IRS1	ORF #3		Y2H(bp)
ISL1	CRK (ORF #3)	No	Y2H(pb)
ISL1	PLCG2	No	Y2H(pb)
ITGB1BP1 (ORF #1)	KRIT1	No	Y2H(bp)
ITGB1BP1	PGRMC1	Yes	
ORF #2	ORF #1		Y2H(bp)
ORF #2	ORF #2		CoIP(pos)
JAK3	LNX1	Yes	Y2H(bp), CoIP(neg)
KCTD13	CRK	Yes	
KCTD13	ORF #1		Y2H(pb)

Interaction		pYd	Evidence
KCTD13	ORF #2		Y2H(pb)
KCTD17	CRK (ORF #3)	No	Y2H(pb)
KCTD17	PLCG2	No	Y2H(pb)
KCTD17	RIN3 (ORF #1)	No	Y2H(pb)
KCTD4	FER	Yes	Y2H(pb)
KCTD4	TNS4	No	Y2H(pb)
KIAA0317	CRKL (ORF #1)	Yes	Y2H(pb)
KLF15	ANKS1A	No	Y2H(pb)
KLF15	APPL1 (ORF #2)	Yes	Y2H(pb)
KLF15	CRK (ORF #2)	No	Y2H(pb)
KLF15	PIK3R3 (ORF #2)	No	Y2H(pb)
KLF15	STAT3	No	
KLF15	ORF #1		Y2H(pb)
KLF15	ORF #2		Y2H(pb)
KLHL20	CRK (ORF #2)	Yes	Y2H(pb)
KLRAQ1	SH2D2A	Yes	Y2H(pb), CoIP(neg)
KRIT1	CCM2	No	Y2H(pb)
KRIT1	ITGB1BP1 (ORF #1)	No	Y2H(pb)
KRTAP10-7	STAT3 (ORF #2)	Yes	Y2H(pb)
KRTAP23-1	TENC1	No	Y2H(pb)
KRTAP3-1	TENC1	No	Y2H(pb)
KRTCAP2	SHD	No	Y2H(pb)
LASP1	CRK	Yes	
LASP1	ORF #1		Y2H(pb)
LASP1	ORF #2		Y2H(pb)
LASP1	ORF #3		CoIP(neg)
LASP1	SH2D2A	No	Y2H(pb), CoIP(pos)
LASP1	STAT3 (ORF #2)	Yes	Y2H(pb)
LCK (ORF #2)	DOK2 (ORF #1)	Yes	Y2H(pb)
LCK (ORF #2)	RIN3 (ORF #2)	No	Y2H(pb)
LCK (ORF #2)	STAT3 (ORF #1)	Yes	Y2H(pb)

Interaction		pYd	Evidence
LCP2	FYB	Yes	Y2H(pb)
LDHAL6B	APPL2	No	Y2H(pb)
LDHAL6B	HSH2D (ORF #2)	Yes	Y2H(pb)
LDHAL6B	SOCS4	Yes	Y2H(pb), CoIP(neg)
LECT1	APPL2	No	Y2H(pb)
LECT1	HSH2D (ORF #2)	Yes	Y2H(pb)
LECT1	SOCS4	Yes	Y2H(pb)
LEFTY2	MAPK8IP2	No	Y2H(pb)
LETMD1	HCK	No	Y2H(pb)
LETMD1	SH2D1A	No	Y2H(pb)
LGALS9C	APPL2	Yes	Y2H(pb), CoIP(neg)
LHX8	CRK (ORF #3)	No	Y2H(pb)
LHX8	PLCG2	Yes	Y2H(pb)
LIX1	APPL2	Yes	Y2H(pb), CoIP(neg)
LIX1	DOK4	No	Y2H(pb)
LMX1A	GRB2 (ORF #2)	Yes	Y2H(pb), CoIP(pos)
LNX1	JAK3	Yes	Y2H(pb), CoIP(neg)
LNX1	PTPN11	Yes	Y2H(pb)
LNX1	SH2D2A	Yes	Y2H(pb), CoIP(neg)
LNX1	SOCS6	Yes	Y2H(pb)
LNX2	NUMB (ORF #1)	No	Y2H(pb)
LOC285398	TENC1	No	Y2H(pb)
LOC492311	MAPK8IP2	Yes	Y2H(pb), CoIP(neg)
LRRFIP1	SH2D2A	Yes	Y2H(pb), CoIP(neg)
LSM8	MAPK8IP2	No	Y2H(pb)
LYN (ORF #2)	PTK2 (ORF #1 WT)	No	Y2H(bp)
MAGEC3	APPL1 (ORF #2)	Yes	Y2H(pb)
MAGEC3	CRK (ORF #2)	Yes	Y2H(pb)
MAPK8IP2	ANAPC10	No	Y2H(bp)
MAPK8IP2	ANGPT1	Yes	Y2H(bp), CoIP(pos)
MAPK8IP2	C1QTNF3	No	Y2H(bp)

Interaction		pYd	Evidence
MAPK8IP2	C22orf39	Yes	Y2H(bp), CoIP(neg)
MAPK8IP2	C8orf33	Yes	Y2H(bp), CoIP(neg)
MAPK8IP2	CRP	No	Y2H(bp)
MAPK8IP2	CTNNBL1	No	Y2H(bp)
MAPK8IP2	DUSP18	No	Y2H(bp)
MAPK8IP2	FGF12	No	Y2H(bp)
MAPK8IP2	FGF13	No	Y2H(bp)
MAPK8IP2	GALNT3	No	Y2H(bp)
MAPK8IP2	HLA-C	No	Y2H(bp)
MAPK8IP2	HSPA1A	No	
MAPK8IP2	ORF #1		Y2H(bp)
MAPK8IP2	ORF #2		CoIP(pos)
MAPK8IP2	ING4	No	Y2H(bp)
MAPK8IP2	LEFTY2	No	Y2H(bp)
MAPK8IP2	LOC492311	Yes	Y2H(bp), CoIP(neg)
MAPK8IP2	LSM8	No	Y2H(bp)
MAPK8IP2	MNDA	No	Y2H(bp)
MAPK8IP2	MRE11A	No	Y2H(bp)
MAPK8IP2	MSRB3	No	Y2H(bp)
MAPK8IP2	NUMB (ORF #1)	No	Y2H(bp)
MAPK8IP2	RABGAP1L	No	
MAPK8IP2	ORF #1		Y2H(bp)
MAPK8IP2	ORF #2		CoIP(pos)
MAPK8IP2	RBM4	No	Y2H(bp)
MAPK8IP2	RYBP	No	
MAPK8IP2	ORF #1		Y2H(bp)
MAPK8IP2	ORF #2		Y2H(bp)
MAPK8IP2	SCLT1	No	Y2H(bp)
MAPK8IP2	SH2D1B	No	
MAPK8IP2	ORF #1		Y2H(bp)
MAPK8IP2	ORF #3		CoIP(neg)

Interaction		pYd	Evidence
MAPK8IP2	SLC39A13	No	
MAPK8IP2	ORF #1		Y2H(bp)
MAPK8IP2	ORF #2		Y2H(bp)
MAPK8IP2	SLPI	No	Y2H(bp)
MAPK8IP2	SPINK2	No	Y2H(bp)
MAPK8IP2	SPSB3	No	Y2H(bp)
MAPK8IP2	SRP19	No	Y2H(bp)
MAPK8IP2	TARBP2	No	Y2H(bp)
MAPK8IP2	TMEM128	Yes	Y2H(bp), CoIP(neg)
MAPK8IP2	UBD	Yes	Y2H(bp), CoIP(pos)
MAPK8IP2	USP46	No	Y2H(bp)
MAPK8IP2	ZNF670	No	Y2H(bp)
MAPK8IP2	ZNRD1	No	Y2H(bp)
MAPKAPK2	STAT3 (ORF #2)	Yes	Y2H(pb)
MAPRE3	APPL2	Yes	Y2H(pb), CoIP(pos)
MAPRE3	DOK4	No	Y2H(pb)
MATK (ORF #2)	FRS3 (ORF #2)	Yes	Y2H(pb), CoIP(neg)
MIST	ANKRD50	No	Y2H(bp)
MIST	ARAP3	Yes	Y2H(bp)
MIST	BMX	No	
MIST	ORF #2		Y2H(bp)
MIST	ORF #3		Y2H(bp)
MIST	C6orf146	Yes	Y2H(bp)
MIST	COQ9	Yes	Y2H(bp)
MIST	CRK (ORF #1)	No	Y2H(bp)
MIST	CYB561D2	Yes	Y2H(bp)
MIST	GLIS3	Yes	Y2H(bp)
MIST	GRB2 (ORF #1)	No	Y2H(bp)
MIST	GSTCD	No	Y2H(bp)
MIST	NME4	Yes	Y2H(bp)
MIST	P4HA2	No	Y2H(bp)

Interaction		pYd	Evidence
MIST	PACSIN3	No	Y2H(bp)
MIST	PIK3R1	Yes	Y2H(bp)
MIST	PIK3R3	Yes	
MIST	ORF #3		Y2H(bp)
MIST	ORF #5		Y2H(bp)
MIST	PPP1R7	Yes	Y2H(bp)
MIST	PTK7	Yes	Y2H(bp)
MIST	RASSF1	No	Y2H(bp)
MIST	RB1	No	Y2H(bp)
MIST	RPAIN	Yes	Y2H(bp)
MIST	SFMBT1	Yes	Y2H(bp)
MIST	SMARCD1	No	Y2H(bp)
MIST	SOCS6	Yes	Y2H(bp)
MIST	STK36	No	Y2H(bp)
MIST	TXK	No	Y2H(bp)
MIST	WWOX	No	Y2H(bp)
MLL3	APPL2	Yes	Y2H(pb)
MLL4	PIK3R3 (ORF #2)	No	Y2H(pb)
MNDA	CRK	No	
MNDA	ORF #1		Y2H(pb)
MNDA	ORF #2		Y2H(pb)
MNDA	MAPK8IP2	No	Y2H(pb)
MNDA	STAT3	No	
MNDA	ORF #1		Y2H(pb)
MNDA	ORF #2		Y2H(pb)
MPEG1	RIN3 (ORF #2)	No	Y2H(pb)
MPEG1	TENC1	No	Y2H(pb)
MPG	CRK (ORF #3)	Yes	Y2H(pb)
MPP5	SOCS4	Yes	Y2H(pb), CoIP(neg)
MPZL1	DOK7	Yes	Y2H(pb)
MPZL1	STAT3 (ORF #2)	Yes	Y2H(pb)

Interaction		pYd	Evidence
MRE11A	MAPK8IP2	No	Y2H(pb)
MRPS22	BTK (ORF #2)	No	Y2H(pb)
MRPS22	SHD	No	Y2H(pb)
MSRB3	MAPK8IP2	No	Y2H(pb)
MST4	GRB14	Yes	Y2H(pb)
MT2A	RIN3	No	
MT2A	ORF #1		Y2H(pb)
MT2A	ORF #2		Y2H(pb)
MT2A	TENC1	No	Y2H(pb)
MTF2	PIK3R3 (ORF #2)	No	Y2H(pb)
MYBPHL	TENC1	No	Y2H(pb)
MYH7B	TENC1	No	Y2H(pb)
MYLIP	CRK (ORF #2)	Yes	Y2H(pb)
MYLIP	DAPP1 (ORF #2)	Yes	Y2H(pb)
MYOZ1	GRB2 (ORF #2)	Yes	Y2H(pb), CoIP(neg)
MYOZ2	CRK (ORF #2)	Yes	Y2H(pb)
MZF1	CBL	No	Y2H(pb)
NACAD	STAT3 (ORF #2)	No	Y2H(pb)
NARG1L	SH2D2A	No	Y2H(pb)
NCK1 (ORF #2)	OLIG1	Yes	Y2H(bp)
NCK1 (ORF #1)	PIK3R3 (ORF #2)	Yes	Y2H(bp)
NCK1 (ORF #1)	SH2D1A	Yes	Y2H(bp)
NCK2	FYB	Yes	Y2H(pb)
NGEF	HCK	Yes	Y2H(pb)
NGEF	SH2D1A	No	Y2H(pb)
NHLRC2	ANKS1A	No	Y2H(pb)
NHP2	NOS1AP	No	Y2H(pb)
NKAP	NOS1AP	No	
ORF #1	NOS1AP		Y2H(pb)
ORF #2	NOS1AP		Y2H(pb)
NME4	MIST	Yes	Y2H(pb)

Interaction		pYd	Evidence
NME4	TNS4	Yes	Y2H(pb)
NMU	TENC1	No	Y2H(pb)
NOS1AP	C1orf62	No	Y2H(bp)
NOS1AP	CKM	No	Y2H(bp)
NOS1AP	FYN (SH2)	No	Y2H(bp)
NOS1AP	HSPD1 (ORF #1)	No	Y2H(bp)
NOS1AP	NHP2	No	Y2H(bp)
NOS1AP	NKAP	No	
NOS1AP	ORF #1		Y2H(bp)
NOS1AP	ORF #2		Y2H(bp)
NOS1AP	PLAGL2	No	Y2H(bp)
NOS1AP	PNP	No	Y2H(bp)
NOS1AP	RP13-36C9.6	No	Y2H(bp)
NOS1AP	RYBP	No	
NOS1AP	ORF #1		Y2H(bp)
NOS1AP	ORF #2		Y2H(bp)
NT5C	TENC1	No	Y2H(pb)
NUFIP2	CRK (ORF #2)	Yes	Y2H(pb)
NUFIP2	STAT3 (ORF #1)	Yes	Y2H(pb)
NUMB (ORF #2)	FRS3 (ORF #2)	Yes	Y2H(pb), CoIP(neg)
NUMB (ORF #2)	IRS1	Yes	Y2H(pb), CoIP(neg)
NUMB (ORF #1)	LNX2	No	Y2H(bp)
NUMB (ORF #1)	MAPK8IP2	No	Y2H(pb)
NUMB	OLIG1	Yes	
ORF #1	OLIG1		Y2H(bp)
ORF #2	OLIG1		CoIP(neg)
NUMB (ORF #1)	TERF2	No	Y2H(bp)
NXT2	STAT3 (ORF #2)	No	Y2H(pb)
OBSL1	TENC1	No	Y2H(pb)
OFCC1	CRK (ORF #2)	Yes	Y2H(pb)
OFCC1	STAT3 (ORF #2)	No	Y2H(pb)

Interaction		pYd	Evidence
OLIG1	DOK4	No	Y2H(pb), CoIP(neg)
OLIG1	DOK5	No	Y2H(pb)
OLIG1	GRB2 (ORF #2)	No	Y2H(pb), PCA(neg), CoIP(neg)
OLIG1	GRB7	Yes	Y2H(pb), CoIP(pos)
OLIG1	HCK	No	Y2H(pb), CoIP(pos)
OLIG1	HSH2D (ORF #2)	No	Y2H(pb)
OLIG1	NCK1 (ORF #2)	Yes	Y2H(pb)
OLIG1	NUMB	Yes	
OLIG1	ORF #1		Y2H(pb)
OLIG1	ORF #2		CoIP(neg)
OLIG1	PIK3R3 (ORF #2)	No	Y2H(pb), CoIP(pos)
OLIG1	PTPN6	Yes	Y2H(pb), CoIP(neg)
OLIG1	RASA1	Yes	Y2H(pb)
OLIG1	SH2D1A	No	Y2H(pb)
OLIG1	SH2D1B	No	
OLIG1	ORF #2		Y2H(pb)
OLIG1	ORF #3		CoIP(pos)
OLIG1	SH2D4A	No	Y2H(pb)
OLIG1	SHD	No	Y2H(pb)
OLIG1	STAT5A	Yes	Y2H(pb), CoIP(neg)
OLIG1	YES1	Yes	Y2H(pb), CoIP(pos)
OSAP	DAPP1 (ORF #2)	No	Y2H(pb)
OSBPL6	SHD	No	Y2H(pb)
OTUD7B	PIK3R3 (ORF #2)	Yes	Y2H(pb), CoIP(pos)
P4HA2	MIST	No	Y2H(pb)
P4HA2	TNS4	Yes	Y2H(pb)
PACRGL	GRB2 (ORF #2)	Yes	Y2H(pb), CoIP(neg)
PACRGL	PIK3R3	No	
PACRGL	ORF #1		CoIP(neg)
PACRGL	ORF #2		Y2H(pb)
PACSIN3	MIST	No	Y2H(pb)

Interaction		pYd	Evidence
PAFAH1B2	CRK (ORF #2)	Yes	Y2H(pb)
PAFAH1B2	STAT3 (ORF #2)	No	Y2H(pb)
PAQR7	STAT3 (ORF #2)	Yes	Y2H(pb)
PARD6A	GRB2 (ORF #2)	Yes	Y2H(pb), PCA(pos), CoIP(pos)
PARD6A	PIK3R3 (ORF #2)	Yes	Y2H(pb), CoIP(pos)
PCDHB5	GRB2 (ORF #2)	Yes	Y2H(pb)
PCDHB5	PIK3R3 (ORF #2)	Yes	Y2H(pb)
PDPK1 (ORF #1)	APBB3	Yes	Y2H(pb), CoIP(neg)
PDPK1 (ORF #2)	CSK	Yes	Y2H(pb), CoIP(neg)
PELI1	IRS1	No	Y2H(pb)
PELI3	PIK3R3 (ORF #2)	Yes	Y2H(pb), CoIP(pos)
PELO	PIK3R3 (ORF #2)	Yes	Y2H(pb)
PER1	TENC1	No	Y2H(pb)
PGRMC1	ITGB1BP1	Yes	
ORF #1	ORF #2		Y2H(pb)
ORF #2	ORF #2		CoIP(pos)
PHC2	CRK (ORF #2)	Yes	Y2H(pb)
PIK3CA	PIK3R3 (ORF #2)	No	Y2H(pb)
PIK3CB	PIK3R3 (ORF #2)	No	Y2H(pb)
PIK3R1	APPL1 (ORF #2)	Yes	Y2H(pb)
PIK3R1	CBL	Yes	Y2H(pb)
PIK3R1	CRK	Yes	
PIK3R1	ORF #1		Y2H(pb)
PIK3R1	ORF #2		Y2H(pb)
PIK3R1	ORF #3		CoIP(neg)
PIK3R1	FER	No	Y2H(pb)
PIK3R1	IRS1	Yes	Y2H(pb), CoIP(pos)
PIK3R1	MIST	Yes	Y2H(pb)
PIK3R1	SH2D2A	Yes	Y2H(pb), PCA(neg), CoIP(pos)
PIK3R1	SOCS1	Yes	Y2H(pb)

Interaction		pYd	Evidence
PIK3R1	STAT3 (ORF #2)	Yes	Y2H(pb)
PIK3R1	TNS4	Yes	Y2H(pb)
PIK3R2	APPL1 (ORF #1)	No	Y2H(pb), CoIP(neg)
PIK3R2	CBL	Yes	Y2H(bp), CoIP(neg)
PIK3R2	CRK	Yes	
PIK3R2	ORF #2		Y2H(pb)
PIK3R2	ORF #3		CoIP(neg)
PIK3R2	STAT3 (ORF #1)	Yes	Y2H(pb)
PIK3R3 (ORF #2)	ABL2 (WT)	Yes	Y2H(pb)
PIK3R3 (ORF #5)	ANKS1A	Yes	Y2H(pb)
PIK3R3	ARL6IP4	Yes	
ORF #1	ARL6IP4		PCA(pos)
ORF #2	ARL6IP4		Y2H(bp)
ORF #2	ARL6IP4		CoIP(pos)
PIK3R3 (ORF #2)	BTK (ORF #2)	Yes	Y2H(bp)
PIK3R3	C10orf81	Yes	
ORF #1	C10orf81		PCA(pos)
ORF #2	C10orf81		Y2H(bp)
ORF #2	C10orf81		CoIP(pos)
PIK3R3 (ORF #2)	C17orf53	No	Y2H(bp)
PIK3R3	C1orf135	No	
ORF #1	C1orf135		CoIP(neg)
ORF #2	C1orf135		Y2H(bp)
PIK3R3 (ORF #2)	C3orf10	No	Y2H(bp)
PIK3R3 (ORF #2)	C3orf34	No	Y2H(bp)
PIK3R3 (ORF #2)	CAMK1	No	Y2H(bp)
PIK3R3	CBL	Yes	
ORF #1	CBL		Y2H(pb)
ORF #2	CBL		Y2H(bp), Y2H(pb), CoIP(pos)
PIK3R3 (ORF #2)	CCDC14	Yes	Y2H(bp)
PIK3R3 (ORF #2)	CCDC33	No	Y2H(bp)

Interaction		pYd	Evidence
PIK3R3 (ORF #2)	CRBN	No	Y2H(bp)
PIK3R3	CRK	Yes	
ORF #3	ORF #2		Y2H(pb)
ORF #4	ORF #2		Y2H(pb)
ORF #5	ORF #2		Y2H(pb)
ORF #3	ORF #1		Y2H(pb)
ORF #5	ORF #1		Y2H(pb)
PIK3R3 (ORF #2)	CRYBA2	Yes	Y2H(bp), CoIP(pos)
PIK3R3 (ORF #2)	E2F6	Yes	Y2H(bp)
PIK3R3 (ORF #2)	EVI1	Yes	Y2H(bp), CoIP(neg)
PIK3R3 (ORF #5)	FER	No	Y2H(pb)
PIK3R3 (ORF #2)	FOXO1	No	Y2H(bp)
PIK3R3 (ORF #2)	FRS3 (ORF #2)	Yes	Y2H(pb), CoIP(pos)
PIK3R3 (ORF #2)	FYN (ORF #1 WT)	No	Y2H(bp)
PIK3R3	HCK	Yes	
ORF #1	HCK		Y2H(pb)
ORF #2	HCK		Y2H(pb), CoIP(pos)
ORF #3	HCK		Y2H(pb)
PIK3R3	HSH2D	Yes	
ORF #1	ORF #2		Y2H(pb)
ORF #2	ORF #2		Y2H(pb), CoIP(neg)
ORF #3	ORF #2		Y2H(pb)
PIK3R3 (ORF #2)	IQUB	Yes	Y2H(bp)
PIK3R3	IRS1	Yes	
ORF #1	IRS1		Y2H(pb)
ORF #2	IRS1		Y2H(pb), CoIP(pos)
ORF #3	IRS1		Y2H(pb)
PIK3R3 (ORF #2)	KLF15	No	Y2H(bp)
PIK3R3	MIST	Yes	
ORF #3	MIST		Y2H(pb)
ORF #5	MIST		Y2H(pb)

Interaction		pYd	Evidence
PIK3R3 (ORF #2)	MLL4	No	Y2H(bp)
PIK3R3 (ORF #2)	MTF2	No	Y2H(bp)
PIK3R3 (ORF #2)	NCK1 (ORF #1)	Yes	Y2H(pb)
PIK3R3 (ORF #2)	OLIG1	No	Y2H(bp), CoIP(pos)
PIK3R3 (ORF #2)	OTUD7B	Yes	Y2H(bp), CoIP(pos)
PIK3R3	PACRGL	No	
ORF #1	PACRGL		CoIP(neg)
ORF #2	PACRGL		Y2H(bp)
PIK3R3 (ORF #2)	PARD6A	Yes	Y2H(bp), CoIP(pos)
PIK3R3 (ORF #2)	PCDHB5	Yes	Y2H(bp)
PIK3R3 (ORF #2)	PELI3	Yes	Y2H(bp), CoIP(pos)
PIK3R3 (ORF #2)	PELO	Yes	Y2H(bp)
PIK3R3 (ORF #2)	PIK3CA	No	Y2H(bp)
PIK3R3 (ORF #2)	PIK3CB	No	Y2H(bp)
PIK3R3 (ORF #2)	PLB1	Yes	Y2H(bp)
PIK3R3 (ORF #2)	PPARA	Yes	Y2H(bp), CoIP(neg)
PIK3R3 (ORF #2)	PPP1R12B	No	Y2H(bp), CoIP(pos)
PIK3R3 (ORF #2)	PTK2 (ORF #1 WT)	No	Y2H(bp), CoIP(pos)
PIK3R3	RBP7	Yes	
ORF #2	RBP7		Y2H(bp), CoIP(pos)
ORF #4	RBP7		PCA(pos)
PIK3R3	SH2D2A	Yes	
ORF #1	SH2D2A		PCA(neg)
ORF #2	SH2D2A		Y2H(pb), CoIP(pos)
PIK3R3	SOCS4	Yes	
ORF #2	SOCS4		Y2H(bp), CoIP(pos)
ORF #4	SOCS4		PCA(pos)
PIK3R3 (ORF #2)	SOCS6	No	Y2H(bp)
PIK3R3 (ORF #2)	SPSB2	Yes	Y2H(bp), CoIP(pos)
PIK3R3	SRC	Yes	
ORF #1	ORF #1		Y2H(pb)

Interaction		pYd	Evidence
ORF #2	ORF #1		CoIP(neg)
ORF #3	ORF #1		Y2H(pb)
PIK3R3	STAT3	Yes	
ORF #3	ORF #2		Y2H(pb)
ORF #5	ORF #1		Y2H(pb)
PIK3R3 (ORF #2)	TERF2	No	Y2H(bp)
PIK3R3 (ORF #2)	TNNC2	No	Y2H(bp)
PIK3R3 (ORF #2)	USP2	Yes	Y2H(bp)
PIK3R3	VCP	Yes	
ORF #2	ORF #1		Y2H(bp)
ORF #2	ORF #2		CoIP(neg)
PIK3R3 (ORF #2)	WBSCR27	Yes	Y2H(bp), CoIP(pos)
PIK3R3	WDFY3	No	
ORF #1	WDFY3		CoIP(pos)
ORF #2	WDFY3		Y2H(bp)
PIK3R3	WDR42A	No	
ORF #2	ORF #1		Y2H(bp)
ORF #2	ORF #2		Y2H(bp)
PIK3R3 (ORF #2)	WRNIP1	No	Y2H(bp)
PIK3R3 (ORF #2)	ZMAT1	No	Y2H(bp)
PIK3R3 (ORF #2)	ZNF281	Yes	Y2H(bp), CoIP(neg)
PIK3R3 (ORF #2)	ZNF451	No	Y2H(bp)
PIK3R3 (ORF #2)	ZNF767	Yes	Y2H(bp), CoIP(pos)
PINK1	APPL2	No	Y2H(pb)
PINK1	HSH2D (ORF #2)	Yes	Y2H(pb)
PINK1	SOCS4	Yes	Y2H(pb), CoIP(pos)
PINK1	STAT3	No	
PINK1	ORF #1		Y2H(pb)
PINK1	ORF #2		Y2H(pb)
PLAGL2	DAPP1 (ORF #1)	No	Y2H(pb), CoIP(neg)
PLAGL2	NOS1AP	No	Y2H(pb)

Interaction		pYd	Evidence
PLB1	PIK3R3 (ORF #2)	Yes	Y2H(pb)
PLB1	RIN3	No	
PLB1	ORF #1		Y2H(pb)
PLB1	ORF #2		Y2H(pb)
PLB1	TENC1	No	Y2H(pb)
PLCG2	ASB9	No	Y2H(bp), CoIP(neg)
PLCG2	C22orf29	No	Y2H(bp)
PLCG2	C5orf35	No	Y2H(bp)
PLCG2	ELK1	Yes	Y2H(bp), CoIP(neg)
PLCG2	EPYC (ORF #1)	Yes	Y2H(bp)
PLCG2	ESD	No	Y2H(bp)
PLCG2	GABPB2 (ORF #1)	No	Y2H(bp)
PLCG2	HSPD1 (ORF #2)	Yes	Y2H(bp)
PLCG2	ISL1	No	Y2H(bp)
PLCG2	KCTD17	No	Y2H(bp)
PLCG2	LHX8	Yes	Y2H(bp)
PLCG2	PSMD3	No	Y2H(bp), CoIP(pos)
PLCG2	PTTG1	Yes	Y2H(bp)
PLCG2	RAD54B	No	Y2H(bp), CoIP(pos)
PLCG2	RALBP1	No	Y2H(bp), CoIP(neg)
PLCG2	TP53	Yes	Y2H(bp)
PLCG2	TWIST2	No	Y2H(bp), CoIP(neg)
PLCG2	ZNF167	No	Y2H(bp)
PLEKHB1	SH2D2A	Yes	Y2H(pb), CoIP(pos)
PNP	NOS1AP	No	Y2H(pb)
POLR3F	RIN3	No	
POLR3F	ORF #1		Y2H(pb)
POLR3F	ORF #2		Y2H(pb)
POLR3F	TENC1	No	Y2H(pb)
PPAPDC2	HSH2D (ORF #2)	Yes	Y2H(pb)
PPAPDC2	SOCS4	Yes	Y2H(pb)

Interaction		pYd	Evidence
PPARA	GRB2 (ORF #2)	Yes	Y2H(pb), CoIP(pos)
PPARA	PIK3R3 (ORF #2)	Yes	Y2H(pb), CoIP(neg)
PPARD	STAT3 (ORF #2)	Yes	Y2H(pb)
PPP1R12B	PIK3R3 (ORF #2)	No	Y2H(pb), CoIP(pos)
PPP1R7	FER	No	Y2H(pb)
PPP1R7	MIST	Yes	Y2H(pb)
PRKACA	CRK (ORF #2)	Yes	Y2H(pb)
PRKCA	CBL	No	Y2H(pb)
PRRG2	CRK (ORF #3)	Yes	Y2H(pb)
PSG11	SHD	Yes	Y2H(pb)
PSMC1	CRK (ORF #2)	No	Y2H(pb)
PSMD3	PLCG2	No	Y2H(pb), CoIP(pos)
PSMD9	RIN3 (ORF #2)	No	Y2H(pb)
PTK2 (ORF #1 KD)	CRK (ORF #1)	Yes	Y2H(pb)
PTK2 (ORF #1 WT)	DOK4	Yes	Y2H(pb), CoIP(neg)
PTK2 (ORF #1 WT)	GRB7	Yes	Y2H(pb), CoIP(pos)
PTK2 (ORF #1 WT)	HCK	No	Y2H(pb), CoIP(pos)
PTK2 (ORF #1 WT)	LYN (ORF #2)	No	Y2H(pb)
PTK2 (ORF #1 WT)	PIK3R3 (ORF #2)	No	Y2H(pb), CoIP(pos)
PTK2 (ORF #1 WT)	PTPN6	Yes	Y2H(pb), CoIP(pos)
PTK2 (ORF #1 WT)	SH2D1A	Yes	Y2H(pb)
PTK2	SH2D1B	No	
ORF #1 WT	ORF #2		Y2H(pb)
ORF #1 WT	ORF #3		CoIP(neg)
PTK2	SH2D2A	Yes	
ORF #1 WT	SH2D2A		Y2H(pb), PCA(neg), CoIP(pos)
ORF #2	SH2D2A		Y2H(pb)
PTK2	SRC	No	
ORF #1 WT	ORF #1		Y2H(pb), CoIP(neg)
ORF #1 WT	ORF #2		Y2H(pb)
ORF #2	ORF #2		Y2H(pb)

Interaction		pYd	Evidence
PTK7	MIST	Yes	Y2H(pb)
PTPN11	LNX1	Yes	Y2H(bp)
PTPN6	OLIG1	Yes	Y2H(bp), CoIP(neg)
PTPN6	PTK2 (ORF #1 WT)	Yes	Y2H(bp), CoIP(pos)
PTTG1	CRK (ORF #3)	No	Y2H(pb)
PTTG1	PLCG2	Yes	Y2H(pb)
RAB2B	CRK (ORF #2)	Yes	Y2H(pb)
RABGAP1	STAT3 (ORF #2)	Yes	Y2H(pb)
RABGAP1L (ORF #1)	ACAP1	Yes	Y2H(bp)
RABGAP1L (ORF #1)	ARFGAP1	Yes	Y2H(bp), CoIP(neg)
RABGAP1L (ORF #1)	C6orf125	Yes	Y2H(bp)
RABGAP1L (ORF #1)	C9orf43	Yes	Y2H(bp), CoIP(pos)
RABGAP1L (ORF #1)	DBN1	Yes	Y2H(bp), CoIP(neg)
RABGAP1L	EVI1	Yes	
ORF #1	EVI1		Y2H(bp)
ORF #3	EVI1		CoIP(pos)
RABGAP1L	MAPK8IP2	No	
ORF #1	MAPK8IP2		CoIP(pos)
ORF #2	MAPK8IP2		Y2H(pb)
RABGAP1L (ORF #1)	RB1	Yes	Y2H(bp), CoIP(neg)
RABGAP1L (ORF #1)	TSPAN2	Yes	Y2H(bp), CoIP(neg)
RABGAP1L (ORF #1)	WDFY3	Yes	Y2H(bp), CoIP(neg)
RAD54B	PLCG2	No	Y2H(pb), CoIP(pos)
RAD54B	SH2D2A	Yes	Y2H(pb), CoIP(pos)
RAI2	APPL2	No	Y2H(pb)
RAI2	DOK5	No	Y2H(pb)
RALBP1	PLCG2	No	Y2H(pb), CoIP(neg)
RANBP3	ANKS1A	No	Y2H(pb)
RANBP3	CHN2	No	Y2H(pb)
RASA1	OLIG1	Yes	Y2H(bp)
RASSF1	MIST	No	Y2H(pb)

Interaction		pYd	Evidence
RB1	ANKS1A	No	Y2H(pb)
RB1	CHN2	No	Y2H(pb)
RB1	GRB2 (ORF #2)	Yes	Y2H(pb), PCA(pos), CoIP(pos)
RB1	MIST	No	Y2H(pb)
RB1	RABGAP1L (ORF #1)	Yes	Y2H(pb), CoIP(neg)
RBM4	MAPK8IP2	No	Y2H(pb)
RBP7	APPL2	No	Y2H(pb)
RBP7	DAPP1 (ORF #1)	No	Y2H(pb)
RBP7	DOK4	No	Y2H(pb)
RBP7	HSH2D (ORF #2)	Yes	Y2H(pb)
RBP7	PIK3R3	Yes	
RBP7	ORF #2		Y2H(pb), CoIP(pos)
RBP7	ORF #4		PCA(pos)
RBP7	SOCS4	Yes	Y2H(pb), PCA(pos), CoIP(pos)
RCAN3	EPS8	No	Y2H(pb), CoIP(neg)
RCAN3	GRB10	Yes	Y2H(pb), CoIP(pos)
RIN1	ANKS1B	No	Y2H(bp)
RIN3	ASB3	No	
ORF #1	ASB3		Y2H(bp)
ORF #2	ASB3		Y2H(bp)
RIN3 (ORF #1)	ASB8	No	Y2H(bp)
RIN3 (ORF #3)	CRK (ORF #4)	No	Y2H(bp)
RIN3 (ORF #3)	CRKL (ORF #1)	No	Y2H(bp)
RIN3	CSRP3	No	
ORF #1	CSRP3		Y2H(bp)
ORF #2	CSRP3		Y2H(bp)
RIN3 (ORF #1)	CYP2C8	No	Y2H(bp)
RIN3 (ORF #1)	CYP46A1	No	Y2H(bp)
RIN3 (ORF #2)	DOK3 (ORF #1)	No	Y2H(bp)

Interaction		pYd	Evidence
RIN3	EPS8L2	No	
ORF #1	EPS8L2		Y2H(bp)
ORF #2	EPS8L2		Y2H(bp)
RIN3 (ORF #2)	EZH2	No	Y2H(bp)
RIN3 (ORF #1)	FAM127C	No	Y2H(bp)
RIN3	FAM46A	No	
ORF #1	FAM46A		Y2H(bp)
ORF #2	FAM46A		Y2H(bp)
RIN3 (ORF #2)	GATA1	No	Y2H(bp)
RIN3 (ORF #3)	GRAP2 (ORF #2)	No	Y2H(bp)
RIN3	ID2	No	
ORF #1	ID2		Y2H(bp)
ORF #2	ID2		Y2H(bp)
RIN3 (ORF #1)	KCTD17	No	Y2H(bp)
RIN3 (ORF #2)	LCK (ORF #2)	No	Y2H(bp)
RIN3 (ORF #2)	MPEG1	No	Y2H(bp)
RIN3	MT2A	No	
ORF #1	MT2A		Y2H(bp)
ORF #2	MT2A		Y2H(bp)
RIN3	PLB1	No	
ORF #1	PLB1		Y2H(bp)
ORF #2	PLB1		Y2H(bp)
RIN3	POLR3F	No	
ORF #1	POLR3F		Y2H(bp)
ORF #2	POLR3F		Y2H(bp)
RIN3 (ORF #2)	PSMD9	No	Y2H(bp)
RIN3 (ORF #2)	SPATA12	No	Y2H(bp)
RIN3	SPRY4	No	
ORF #1	SPRY4		Y2H(bp)
ORF #2	SPRY4		Y2H(bp)
RIN3 (ORF #2)	SUV39H1	No	Y2H(bp)

Interaction		pYd	Evidence
RIN3 (ORF #1)	SYMPK	No	Y2H(bp)
RIN3 (ORF #1)	TGM1	No	Y2H(bp)
RIN3	TM4SF19	No	
ORF #1	TM4SF19		Y2H(bp)
ORF #2	TM4SF19		Y2H(bp)
RIN3 (ORF #2)	TSC1 (ORF #2)	No	Y2H(bp)
RIN3 (ORF #1)	TSSK3	No	Y2H(bp)
RIN3 (ORF #1)	ZBTB24	No	Y2H(bp)
RIN3	ZNF639	No	
ORF #1	ZNF639		Y2H(bp)
ORF #2	ZNF639		Y2H(bp)
RIN3	ZNF655	No	
ORF #1	ZNF655		Y2H(bp)
ORF #2	ZNF655		Y2H(bp)
RIN3 (ORF #1)	ZNF71	No	Y2H(bp)
RP13-36C9.6	NOS1AP	No	Y2H(pb)
RPAIN	MIST	Yes	Y2H(pb)
RPAIN	TNS4	Yes	Y2H(pb)
RPS6	FRS2	No	Y2H(pb)
RTDR1	EPS8	No	Y2H(pb)
RYBP	CRK	No	
ORF #2	ORF #1		Y2H(pb)
ORF #2	ORF #2		Y2H(pb)
RYBP	MAPK8IP2	No	
ORF #1	MAPK8IP2		Y2H(pb)
ORF #2	MAPK8IP2		Y2H(pb)
RYBP	NOS1AP	No	
ORF #1	NOS1AP		Y2H(pb)
ORF #2	NOS1AP		Y2H(pb)
SCAPER	APPL1 (ORF #1)	No	Y2H(pb)
SCLT1	MAPK8IP2	No	Y2H(pb)

Interaction		pYd	Evidence
SCOC	DOK5	Yes	Y2H(pb), CoIP(neg)
SDCCAG8	TENC1	No	Y2H(pb)
SEMA4D	CRK (ORF #3)	Yes	Y2H(pb), CoIP(neg)
SEPT6	CRK	No	
ORF #1	ORF #1		Y2H(pb)
ORF #2	ORF #1		Y2H(pb)
ORF #2	ORF #2		Y2H(pb)
SERPINA5	CBL	No	Y2H(pb)
SERPINA5	EPS8	No	Y2H(pb)
SFMBT1	MIST	Yes	Y2H(pb)
SH2B1	CSAD	Yes	Y2H(bp)
SH2B1	FRS3 (ORF #2)	No	Y2H(pb)
SH2B3	TENC1	No	Y2H(pb)
SH2D1A	ATF3	Yes	Y2H(bp)
SH2D1A	CCDC74A	No	Y2H(bp)
SH2D1A	LETMD1	No	Y2H(bp)
SH2D1A	NCK1 (ORF #1)	Yes	Y2H(pb)
SH2D1A	NGEF	No	Y2H(bp)
SH2D1A	OLIG1	No	Y2H(bp)
SH2D1A	PTK2 (ORF #1 WT)	Yes	Y2H(bp)
SH2D1A	ZNHIT1	Yes	Y2H(bp)
SH2D1B	HSPA1A	Yes	
ORF #3	ORF #1		Y2H(bp)
ORF #3	ORF #2		CoIP(pos)
SH2D1B	MAPK8IP2	No	
ORF #1	MAPK8IP2		Y2H(pb)
ORF #3	MAPK8IP2		CoIP(neg)
SH2D1B	OLIG1	No	
ORF #2	OLIG1		Y2H(bp)
ORF #3	OLIG1		CoIP(pos)

Interaction		pYd	Evidence
SH2D1B	PTK2	No	
ORF #2	ORF #1 WT		Y2H(bp)
ORF #3	ORF #1 WT		CoIP(neg)
SH2D1B (ORF #3)	STAT4	Yes	Y2H(bp), CoIP(neg)
SH2D2A	APPL1	Yes	
SH2D2A	ORF #1		CoIP(neg)
SH2D2A	ORF #2		Y2H(pb)
SH2D2A	ARID5A	No	Y2H(bp)
SH2D2A	ASB3	Yes	Y2H(bp), CoIP(pos)
SH2D2A	C17orf82	No	Y2H(bp)
SH2D2A	CRK	Yes	
SH2D2A	ORF #1		Y2H(pb)
SH2D2A	ORF #2		Y2H(pb)
SH2D2A	ORF #3		CoIP(neg)
SH2D2A	FAM46A	Yes	Y2H(bp), CoIP(pos)
SH2D2A	FAM46B	Yes	Y2H(bp), CoIP(pos)
SH2D2A	KLRAQ1	Yes	Y2H(bp), CoIP(neg)
SH2D2A	LASP1	No	Y2H(bp), CoIP(pos)
SH2D2A	LNX1	Yes	Y2H(bp), CoIP(neg)
SH2D2A	LRRFIP1	Yes	Y2H(bp), CoIP(neg)
SH2D2A	NARG1L	No	Y2H(bp)
SH2D2A	PIK3R1	Yes	Y2H(bp), PCA(neg), CoIP(pos)
SH2D2A	PIK3R3	Yes	
SH2D2A	ORF #1		PCA(neg)
SH2D2A	ORF #2		Y2H(bp), CoIP(pos)
SH2D2A	PLEKHB1	Yes	Y2H(bp), CoIP(pos)
SH2D2A	PTK2	Yes	
SH2D2A	ORF #1 WT		Y2H(bp), PCA(neg), CoIP(pos)
SH2D2A	ORF #2		Y2H(bp)
SH2D2A	RAD54B	Yes	Y2H(bp), CoIP(pos)

Interaction		pYd	Evidence
SH2D2A	SOCS3	Yes	Y2H(pb)
SH2D2A	SPTB	Yes	Y2H(bp)
SH2D2A	STAT3	Yes	
SH2D2A	ORF #1		Y2H(pb)
SH2D2A	ORF #2		Y2H(pb)
SH2D2A	TSC1	Yes	
SH2D2A	ORF #1		Y2H(bp)
SH2D2A	ORF #2		Y2H(bp), CoIP(neg)
SH2D2A	WDR40B	Yes	Y2H(bp), CoIP(pos)
SH2D2A	ZHX3	Yes	Y2H(bp), CoIP(neg)
SH2D2A	ZSCAN1	Yes	Y2H(bp)
SH2D3A	CENPB	No	Y2H(bp)
SH2D4A	OLIG1	No	Y2H(bp)
SH3BP2 (ORF #1)	ARHGAP10	Yes	Y2H(bp)
SH3BP2 (ORF #2)	STAT3 (ORF #2)	Yes	Y2H(pb)
SHD	BECN1	No	Y2H(bp)
SHD	CAPN10	Yes	Y2H(bp)
SHD	KRTCAP2	No	Y2H(bp)
SHD	MRPS22	No	Y2H(bp)
SHD	OLIG1	No	Y2H(bp)
SHD	OSBPL6	No	Y2H(bp)
SHD	PSG11	Yes	Y2H(bp)
SHE	DOK3 (ORF #1)	Yes	Y2H(bp)
SHE	TRIT1	Yes	Y2H(bp)
SLC39A13	MAPK8IP2	No	
ORF #1	MAPK8IP2		Y2H(pb)
ORF #2	MAPK8IP2		Y2H(pb)
SLPI	MAPK8IP2	No	Y2H(pb)
SMARCD1	ANKS1A	No	Y2H(pb)
SMARCD1	CHN2	No	Y2H(pb)
SMARCD1	MIST	No	Y2H(pb)

Interaction		pYd	Evidence
SOCS1	CRK (ORF #2)	Yes	Y2H(pb)
SOCS1	CRKL (ORF #1)	Yes	Y2H(pb)
SOCS1	PIK3R1	Yes	Y2H(bp)
SOCS3	SH2D2A	Yes	Y2H(bp)
SOCS3	TRDN	Yes	Y2H(bp)
SOCS3	YWHAQ	No	Y2H(bp)
SOCS4	FRS3 (ORF #2)	No	Y2H(pb), CoIP(pos)
SOCS4	GRB2 (ORF #2)	Yes	Y2H(pb), PCA(neg), CoIP(pos)
SOCS4	LDHAL6B	Yes	Y2H(bp), CoIP(neg)
SOCS4	LECT1	Yes	Y2H(bp)
SOCS4	MPP5	Yes	Y2H(bp), CoIP(neg)
SOCS4	PIK3R3	Yes	
SOCS4	ORF #2		Y2H(pb), CoIP(pos)
SOCS4	ORF #4		PCA(pos)
SOCS4	PINK1	Yes	Y2H(bp), CoIP(pos)
SOCS4	PPAPDC2	Yes	Y2H(bp)
SOCS4	RBP7	Yes	Y2H(bp), PCA(pos), CoIP(pos)
SOCS4	TNK2	Yes	Y2H(bp)
SOCS6	APPL1 (ORF #2)	Yes	Y2H(pb)
SOCS6	CRK (ORF #2)	Yes	Y2H(pb)
SOCS6	FRS3 (ORF #2)	No	Y2H(pb)
SOCS6	LNX1	Yes	Y2H(bp)
SOCS6	MIST	Yes	Y2H(pb)
SOCS6	PIK3R3 (ORF #2)	No	Y2H(pb)
SPATA12	RIN3 (ORF #2)	No	Y2H(pb)
SPATA12	TENC1	No	Y2H(pb)
SPINK2	MAPK8IP2	No	Y2H(pb)
SPRY2	TENC1	No	Y2H(pb)
SPRY4	RIN3	No	
SPRY4	ORF #1		Y2H(pb)

Interaction		pYd	Evidence
SPRY4	ORF #2		Y2H(pb)
SPRY4	TENC1	No	Y2H(pb)
SPSB2	PIK3R3 (ORF #2)	Yes	Y2H(pb), CoIP(pos)
SPSB3	MAPK8IP2	No	Y2H(pb)
SPTB	SH2D2A	Yes	Y2H(pb)
SPTB	STAT1	Yes	Y2H(pb)
SRC	PIK3R3	Yes	
ORF #1	ORF #1		Y2H(bp)
ORF #1	ORF #2		CoIP(neg)
ORF #1	ORF #3		Y2H(bp)
SRC	PTK2	No	
ORF #1	ORF #1 WT		Y2H(bp), CoIP(neg)
ORF #2	ORF #1 WT		Y2H(bp)
ORF #2	ORF #2		Y2H(bp)
SRP19	MAPK8IP2	No	Y2H(pb)
STAT1	SPTB	Yes	Y2H(bp)
STAT2	DOK4	Yes	Y2H(pb), CoIP(neg)
STAT3 (ORF #1)	ABL2 (KD)	Yes	Y2H(bp)
STAT3 (ORF #2)	ATF3	No	Y2H(bp)
STAT3 (ORF #2)	BATF3	No	Y2H(bp)
STAT3 (ORF #2)	BMX (ORF #2)	No	Y2H(bp)
STAT3 (ORF #1)	BRWD1	Yes	Y2H(bp)
STAT3	C10orf18	Yes	
ORF #1	C10orf18		Y2H(bp)
ORF #2	C10orf18		Y2H(bp)
STAT3 (ORF #2)	CA8	Yes	Y2H(bp)
STAT3 (ORF #2)	CBL	Yes	Y2H(bp)
STAT3 (ORF #2)	CCDC87	Yes	Y2H(bp)
STAT3	CHTF18	No	
ORF #1	CHTF18		Y2H(bp)
ORF #2	CHTF18		Y2H(bp)

Interaction		pYd	Evidence
STAT3	CNDP2	No	
ORF #1	CNDP2		Y2H(bp)
ORF #2	CNDP2		Y2H(bp)
STAT3 (ORF #1)	DOK2 (ORF #2)	Yes	Y2H(bp)
STAT3	DOK3	Yes	
ORF #1	ORF #1		Y2H(bp)
ORF #1	ORF #2		Y2H(bp)
ORF #2	ORF #2		Y2H(bp)
STAT3	DTNA	Yes	
ORF #1	DTNA		Y2H(bp)
ORF #2	DTNA		Y2H(bp)
STAT3 (ORF #2)	FAM117B	Yes	Y2H(bp)
STAT3	GSTCD	Yes	
ORF #1	GSTCD		Y2H(bp)
ORF #2	GSTCD		Y2H(bp)
STAT3	HESX1	Yes	
ORF #1	HESX1		Y2H(bp)
ORF #2	HESX1		Y2H(bp)
STAT3	KLF15	No	
ORF #1	KLF15		Y2H(bp)
ORF #2	KLF15		Y2H(bp)
STAT3 (ORF #2)	KRTAP10-7	Yes	Y2H(bp)
STAT3 (ORF #2)	LASP1	Yes	Y2H(bp)
STAT3 (ORF #1)	LCK (ORF #2)	Yes	Y2H(bp)
STAT3 (ORF #2)	MAPKAPK2	Yes	Y2H(bp)
STAT3	MNDA	No	
ORF #1	MNDA		Y2H(bp)
ORF #2	MNDA		Y2H(bp)
STAT3 (ORF #2)	MPZL1	Yes	Y2H(bp)
STAT3 (ORF #2)	NACAD	No	Y2H(bp)
STAT3 (ORF #1)	NUFIP2	Yes	Y2H(bp)

Interaction		pYd	Evidence
STAT3 (ORF #2)	NXT2	No	Y2H(bp)
STAT3 (ORF #2)	OFCC1	No	Y2H(bp)
STAT3 (ORF #2)	PAFAH1B2	No	Y2H(bp)
STAT3 (ORF #2)	PAQR7	Yes	Y2H(bp)
STAT3 (ORF #2)	PIK3R1	Yes	Y2H(bp)
STAT3 (ORF #1)	PIK3R2	Yes	Y2H(bp)
STAT3	PIK3R3	Yes	
ORF #1	ORF #5		Y2H(bp)
ORF #2	ORF #3		Y2H(bp)
STAT3	PINK1	No	
ORF #1	PINK1		Y2H(bp)
ORF #2	PINK1		Y2H(bp)
STAT3 (ORF #2)	PPARD	Yes	Y2H(bp)
STAT3 (ORF #2)	RABGAP1	Yes	Y2H(bp)
STAT3	SH2D2A	Yes	
ORF #1	SH2D2A		Y2H(bp)
ORF #2	SH2D2A		Y2H(bp)
STAT3 (ORF #2)	SH3BP2 (ORF #2)	Yes	Y2H(bp)
STAT3 (ORF #2)	TDG	No	Y2H(bp)
STAT3	TM4SF19	Yes	
ORF #1	TM4SF19		Y2H(bp)
ORF #2	TM4SF19		Y2H(bp)
STAT3 (ORF #2)	TWIST1	Yes	Y2H(bp)
STAT3 (ORF #1)	VPS39	Yes	Y2H(bp)
STAT3 (ORF #2)	WDFY3	Yes	Y2H(bp)
STAT3 (ORF #2)	ZNF281	Yes	Y2H(bp)
STAT3 (ORF #2)	ZNF557	No	Y2H(bp)
STAT3 (ORF #2)	ZNF829	Yes	Y2H(bp)
STAT4	CRK (ORF #3)	Yes	Y2H(pb), CoIP(neg)
STAT4	SH2D1B (ORF #3)	Yes	Y2H(pb), CoIP(neg)
STAT5A	CBL	Yes	Y2H(bp), CoIP(neg)

Interaction		pYd	Evidence
STAT5A	OLIG1	Yes	Y2H(bp), CoIP(neg)
STK36	MIST	No	Y2H(pb)
STRN4	CRK (ORF #3)	No	Y2H(pb)
SUV39H1	RIN3 (ORF #2)	No	Y2H(pb)
SUV39H1	TENC1	No	Y2H(pb)
SYK (ORF #1 WT)	FAM46A	Yes	Y2H(bp), CoIP(pos)
SYMPK	RIN3 (ORF #1)	No	Y2H(pb)
TARBP2	MAPK8IP2	No	Y2H(pb)
TCAP	CRK (ORF #2)	Yes	Y2H(pb)
TDG	CRK (ORF #2)	No	Y2H(pb)
TDG	STAT3 (ORF #2)	No	Y2H(pb)
TENC1	ADCK5	No	Y2H(bp)
TENC1	AP1S2	No	Y2H(bp)
TENC1	ASB3	No	Y2H(bp)
TENC1	C19orf66	No	Y2H(bp)
TENC1	C21orf77	No	Y2H(bp)
TENC1	COX6B2	No	Y2H(bp)
TENC1	CRKL (ORF #1)	No	Y2H(bp)
TENC1	DDIT4L	No	Y2H(bp)
TENC1	DOK3 (ORF #1)	No	Y2H(bp)
TENC1	EPB41L4A	No	Y2H(bp)
TENC1	EPS8L2	No	Y2H(bp)
TENC1	FABP7	No	Y2H(bp)
TENC1	FAM114A1	No	Y2H(bp)
TENC1	FAM46A	No	Y2H(bp)
TENC1	FGF21	No	Y2H(bp)
TENC1	FOXO1	No	Y2H(bp)
TENC1	GATA1	No	Y2H(bp)
TENC1	GBA3	No	Y2H(bp)
TENC1	HUWE1	No	Y2H(bp)
TENC1	ID2	No	Y2H(bp)

Interaction		pYd	Evidence
TENC1	KRTAP23-1	No	Y2H(bp)
TENC1	KRTAP3-1	No	Y2H(bp)
TENC1	LOC285398	No	Y2H(bp)
TENC1	MPEG1	No	Y2H(bp)
TENC1	MT2A	No	Y2H(bp)
TENC1	MYBPHL	No	Y2H(bp)
TENC1	MYH7B	No	Y2H(bp)
TENC1	NMU	No	Y2H(bp)
TENC1	NT5C	No	Y2H(bp)
TENC1	OBSL1	No	Y2H(bp)
TENC1	PER1	No	Y2H(bp)
TENC1	PLB1	No	Y2H(bp)
TENC1	POLR3F	No	Y2H(bp)
TENC1	SDCCAG8	No	Y2H(bp)
TENC1	SH2B3	No	Y2H(bp)
TENC1	SPATA12	No	Y2H(bp)
TENC1	SPRY2	No	Y2H(bp)
TENC1	SPRY4	No	Y2H(bp)
TENC1	SUV39H1	No	Y2H(bp)
TENC1	TM4SF19	No	Y2H(bp)
TENC1	TSC1 (ORF #2)	No	Y2H(bp)
TENC1	TWSG1	No	Y2H(bp)
TENC1	WDFY3	No	Y2H(bp)
TENC1	WDR23	No	Y2H(bp)
TENC1	WDR42A (ORF #2)	No	Y2H(bp)
TENC1	ZNF586	No	Y2H(bp)
TENC1	ZNF655	No	Y2H(bp)
TENC1	ZNRD1	No	Y2H(bp)
TERF2	NUMB (ORF #1)	No	Y2H(pb)
TERF2	PIK3R3 (ORF #2)	No	Y2H(pb)
TGM1	RIN3 (ORF #1)	No	Y2H(pb)

Interaction		pYd	Evidence
TM4SF19	CRK (ORF #2)	Yes	Y2H(pb)
TM4SF19	DAPP1 (ORF #2)	No	Y2H(pb)
TM4SF19	RIN3	No	
TM4SF19	ORF #1		Y2H(pb)
TM4SF19	ORF #2		Y2H(pb)
TM4SF19	STAT3	Yes	
TM4SF19	ORF #1		Y2H(pb)
TM4SF19	ORF #2		Y2H(pb)
TM4SF19	TENC1	No	Y2H(pb)
TMEM128	GRB2 (ORF #2)	Yes	Y2H(pb), PCA(neg), CoIP(pos)
TMEM128	MAPK8IP2	Yes	Y2H(pb), CoIP(neg)
TMEM168	CRKL (ORF #1)	Yes	Y2H(pb)
TNFAIP1	DAPP1 (ORF #2)	Yes	Y2H(pb)
TNK2	HSH2D (ORF #2)	No	Y2H(pb)
TNK2	SOCS4	Yes	Y2H(pb)
TNNC2	PIK3R3 (ORF #2)	No	Y2H(pb)
TNS1	FRS3 (ORF #2)	No	Y2H(pb)
TNS4	GSTCD	Yes	Y2H(bp)
TNS4	KCTD4	No	Y2H(bp)
TNS4	NME4	Yes	Y2H(bp)
TNS4	P4HA2	Yes	Y2H(bp)
TNS4	PIK3R1	Yes	Y2H(bp)
TNS4	RPAIN	Yes	Y2H(bp)
TP53	APPL1 (ORF #1)	No	Y2H(pb)
TP53	CRK (ORF #3)	No	Y2H(pb)
TP53	PLCG2	Yes	Y2H(pb)
TP53RK	CHN2	No	Y2H(pb)
TRDN	SOCS3	Yes	Y2H(pb)
TRIT1	ANKS1A	No	Y2H(pb)
TRIT1	SHE	Yes	Y2H(pb)
TSC1 (ORF #1)	APPL2	No	Y2H(pb)

Interaction		pYd	Evidence
TSC1 (ORF #1)	DOK5	No	Y2H(pb)
TSC1 (ORF #1)	HSH2D (ORF #2)	Yes	Y2H(pb)
TSC1 (ORF #2)	RIN3 (ORF #2)	No	Y2H(pb)
TSC1	SH2D2A	Yes	
ORF #1	SH2D2A		Y2H(pb)
ORF #2	SH2D2A		Y2H(pb), CoIP(neg)
TSC1 (ORF #2)	TENC1	No	Y2H(pb)
TSPAN2	GRB2 (ORF #2)	Yes	Y2H(pb), PCA(pos), CoIP(pos)
TSPAN2	RABGAP1L (ORF #1)	Yes	Y2H(pb), CoIP(neg)
TSSK3	RIN3 (ORF #1)	No	Y2H(pb)
TUBA1C	CRK (ORF #3)	No	Y2H(pb)
TWIST1	STAT3 (ORF #2)	Yes	Y2H(pb)
TWIST2	CCM2	Yes	Y2H(pb), CoIP(pos)
TWIST2	CRK (ORF #3)	No	Y2H(pb), CoIP(neg)
TWIST2	PLCG2	No	Y2H(pb), CoIP(neg)
TWSG1	TENC1	No	Y2H(pb)
TXK	MIST	No	Y2H(pb)
UBD	MAPK8IP2	Yes	Y2H(pb), CoIP(pos)
USP2	PIK3R3 (ORF #2)	Yes	Y2H(pb)
USP46	MAPK8IP2	No	Y2H(pb)
VAC14	CRK (ORF #1)	Yes	Y2H(pb)
VASP	FRS3	Yes	
ORF #1	ORF #2		Y2H(pb)
ORF #2	ORF #2		CoIP(neg)
VCP	GRB2	Yes	
ORF #1	ORF #2		Y2H(pb)
ORF #2	ORF #2		CoIP(pos)
VCP	PIK3R3	Yes	
ORF #1	ORF #2		Y2H(pb)
ORF #2	ORF #2		CoIP(neg)
VPS39	STAT3 (ORF #1)	Yes	Y2H(pb)

Interaction		pYd	Evidence
VSTM2L	CCM2	Yes	Y2H(pb), CoIP(pos)
WBSCR27	GRB2 (ORF #2)	No	Y2H(pb), PCA(neg), CoIP(neg)
WBSCR27	PIK3R3 (ORF #2)	Yes	Y2H(pb), CoIP(pos)
WDFY3	GRB2 (ORF #2)	Yes	Y2H(pb), CoIP(neg)
WDFY3	PIK3R3	No	
WDFY3	ORF #2		Y2H(pb)
WDFY3	ORF #1		CoIP(pos)
WDFY3	RABGAP1L (ORF #1)	Yes	Y2H(pb), CoIP(neg)
WDFY3	STAT3 (ORF #2)	Yes	Y2H(pb)
WDFY3	TENC1	No	Y2H(pb)
WDR20	DAPP1 (ORF #1)	Yes	Y2H(pb), CoIP(pos)
WDR23	TENC1	No	Y2H(pb)
WDR40B	SH2D2A	Yes	Y2H(pb), CoIP(pos)
WDR42A	PIK3R3	No	
ORF #1	ORF #2		Y2H(pb)
ORF #2	ORF #2		Y2H(pb)
WDR42A (ORF #2)	TENC1	No	Y2H(pb)
WDR83	CRK	No	
WDR83	ORF #1		Y2H(pb)
WDR83	ORF #2		Y2H(pb)
WIZ	DOK5	No	Y2H(pb)
WRNIP1	PIK3R3 (ORF #2)	No	Y2H(pb)
WWOX	MIST	No	Y2H(pb)
XAGE2XX	EPS8	Yes	Y2H(pb)
YES1	BECN1	Yes	Y2H(bp), CoIP(neg)
YES1	OLIG1	Yes	Y2H(bp), CoIP(pos)
YPEL3	DOK7	No	Y2H(pb)
YWHAQ	SOCS3	No	Y2H(pb)
ZBTB24	RIN3 (ORF #1)	No	Y2H(pb)
ZHX3	SH2D2A	Yes	Y2H(pb), CoIP(neg)
ZMAT1	GRB2 (ORF #2)	No	Y2H(pb)

Interaction		pYd	Evidence
ZMAT1	PIK3R3 (ORF #2)	No	Y2H(pb)
ZNF167	CRK (ORF #3)	No	Y2H(pb)
ZNF167	PLCG2	No	Y2H(pb)
ZNF281	PIK3R3 (ORF #2)	Yes	Y2H(pb), CoIP(neg)
ZNF281	STAT3 (ORF #2)	Yes	Y2H(pb)
ZNF451	PIK3R3 (ORF #2)	No	Y2H(pb)
ZNF557	CRK (ORF #2)	No	Y2H(pb)
ZNF557	STAT3 (ORF #2)	No	Y2H(pb)
ZNF586	TENC1	No	Y2H(pb)
ZNF639	RIN3	No	
ZNF639	ORF #1		Y2H(pb)
ZNF639	ORF #2		Y2H(pb)
ZNF655	RIN3	No	
ZNF655	ORF #1		Y2H(pb)
ZNF655	ORF #2		Y2H(pb)
ZNF655	TENC1	No	Y2H(pb)
ZNF670	MAPK8IP2	No	Y2H(pb)
ZNF71	RIN3 (ORF #1)	No	Y2H(pb)
ZNF767	PIK3R3 (ORF #2)	Yes	Y2H(pb), CoIP(pos)
ZNF829	APPL1 (ORF #2)	Yes	Y2H(pb)
ZNF829	STAT3 (ORF #2)	Yes	Y2H(pb)
ZNHIT1	DAPP1 (ORF #1)	Yes	Y2H(pb)
ZNHIT1	GRB7	Yes	Y2H(pb)
ZNHIT1	SH2D1A	Yes	Y2H(pb)
ZNRD1	MAPK8IP2	No	Y2H(pb)
ZNRD1	TENC1	No	Y2H(pb)
ZSCAN1	SH2D2A	Yes	Y2H(pb)

Appendix C

Screened ORF-kinase combinations

For each ORF screened in Round One and Round Two, the exact number of bait strains with each kinase construct resulting from each parent strain is provided. The numbers of prey genes represented in the matrix were 13,807 and 17,007, respectively. Unsuccessful ORFs, i.e. ORFs without interactions, are marked in *italics*. NLS = nuclear localization signal; U2 = L40ccU2

Round One

Kinase:	ABL2				BMX				FES				FRK				FYN				JAK2				PTK2				SYK				TNK1			
Parent strain:	**U2**		**L40c**		**U2**		**L40c**		**U2**		**L40c**		**U2**		**L40c**		**U2**		**L40c**		**U2**		**L40c**		**U2**		**L40c**		**U2**		**L40c**		**U2**		**L40c**	
NLS:	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
ORF:																																				
ABL2 (WT)	2	2	1	2	2	1	1	0	2	2	2	1	1	1	1	2	2	1	2	2	1	0	1	1	1	2	2	2	0	1	1	0	0	1	0	1
ANKS1B (ORF #2)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
APBB2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
APBB3 (ORF #2)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
APPL2	1	1	1	0	1	1	0	1	2	1	1	1	1	2	1	1	1	1	1	0	1	1	1	0	0	1	1	2	1	0	2	2	1	2	2	1
BCAR3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2
BLK	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
BMX (ORF #1 WT)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	1	2	2
BTK (ORF #2)	2	2	2	2	2	2	2	1	2	2	2	0	2	2	2	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	2
CBL	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
CBLB (ORF #1)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Round One (continued)

Kinase:	ABL2				BMX				FES				FRK				FYN				JAK2				PTK2				SYK				TNK1			
Parent strain:	U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c	
NLS:	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
ORF:																																				
CBLC (ORF #2)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2
CCM2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	1	2
CHN1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
CISH	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
CRK (ORF #3)	0	0	1	2	0	0	1	1	0	0	1	0	0	0	2	2	0	0	1	1	0	0	0	2	0	0	1	1	0	0	1	1	0	0	0	1
CSK	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
DOK3 (ORF #1)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
DOK4	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2
DOK5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
DOK6	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
EPS8	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
FAM43A	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
FES (WT)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2

Round One (continued)

Kinase:	ABL2				BMX				FES				FRK				FYN				JAK2				PTK2				SYK				TNK1			
Parent strain:	U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c	
NLS:	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
ORF:																																				
FRK (WT)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
FRS2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
FRS3 (ORF #1)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2
FYN (ORF #1 WT)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
GRB2 (ORF #2)	2	0	0	0	1	2	0	0	0	1	1	1	2	2	1	2	1	0	0	0	2	2	1	0	2	1	1	0	0	2	1	1	2	2	1	1
GRB7	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
GRB10	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
GRB14	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
HCK	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
HSH2D (ORF #2)	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
IRS1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Round One (continued)

Kinase:	ABL2				BMX				FES				FRK				FYN				JAK2				PTK2				SYK				TNK1			
Parent strain:	U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c	
NLS:	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
ORF:																																				
ITGB1BP1 (ORF #1)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
JAK3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
LCP2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2
LDLRAP1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
LYN (ORF #2)	0	0	2	1	0	0	2	2	0	0	2	2	0	0	2	2	0	0	2	2	0	0	2	2	0	0	2	2	0	0	2	2	0	0	2	2
MAPK8IP2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2
MATK (ORF #2)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
NCK1 (ORF #2)	2	1	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	0	2	2	1	2	2	2	2	1	2	1
NOS1AP	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
NUMB (ORF #1)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
PID1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Round One (continued)

Kinase:	ABL2				BMX				FES				FRK				FYN				JAK2				PTK2				SYK				TNK1			
Parent strain:	U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c	
NLS:	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
ORF:																																				
PIK3R2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
PIK3R3 (ORF #2)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
PLCG2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	1	2	1
PTPN11	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
PTPN6	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
RABGAP1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
RABGAP1L (ORF #1)	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2
RASA1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
RIN1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
SH2B1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2
SH2B3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2
SH2D1A	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
SH2D1B (ORF #2)	2	0	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Round One (continued)

Kinase:	ABL2				BMX				FES				FRK				FYN				JAK2				PTK2				SYK				TNK1			
Parent strain:	U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c	
NLS:	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
ORF:																																				
SH2D2A	2	0	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2
SH2D3A	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
SH2D3C	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
SH2D4A	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
SHC4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
SHD	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	0	2	0
SOCS2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
SOCS4	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2
SOCS5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
SOCS6	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2
SRC (ORF #2)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2
STAP1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
STAT1	2	2	2	1	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
STAT2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Round One (continued)

Kinase:	ABL2				BMX				FES				FRK				FYN				JAK2				PTK2				SYK				TNK1			
Parent strain:	U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c		U2		L40c	
NLS:	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
ORF:																																				
STAT5A	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	1	2	2	1	2	2	2	2	2
STAT6	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
SYK (ORF #1 WT)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
TNS1	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2
TNS3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
TYK2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
VAV1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
VAV3	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
YES1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
ZAP70 (ORF #1)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2

Round Two

Kinase:	ABL2				FYN				TNK1			
Parent strain:	U2		L40c		U2		L40c		U2		L40c	
NLS:	−	+	−	+	−	+	−	+	−	+	−	+
ORF:												
ANKS1A	4	4	4	4	4	4	4	4	4	4	4	4
APBA2	4	4	4	4	4	4	4	4	4	4	4	4
APBB1	4	4	4	4	4	4	4	4	4	4	4	4
APPL1 (ORF #2)	4	4	4	4	4	4	4	4	4	4	4	4
BLNK	4	3	4	4	4	4	4	4	4	4	4	4
CHN2	4	4	4	4	4	4	4	4	4	4	4	4
CRK (ORF #1)	4	4	4	4	4	4	4	4	4	4	4	4
CRK (ORF #2)	4	4	4	4	4	4	4	4	4	4	4	4
DAB2	4	4	4	4	4	4	4	4	4	4	4	4
DOK1	4	4	4	4	4	4	4	4	4	4	4	4
DOK2 (ORF #1)	4	4	4	4	4	4	4	4	4	4	4	4
DOK7	4	4	4	4	4	4	4	4	4	4	4	4
EPS8L2	4	4	4	4	4	4	4	4	4	4	4	4
FAM43B	4	4	4	4	4	4	4	4	4	4	4	4
FER	4	4	4	4	4	4	4	4	4	4	4	4
GRAP	4	4	4	4	4	4	4	4	4	4	4	4
INPP5D (ORF #1)	4	4	4	4	4	4	4	4	4	4	4	4
INPP5D (ORF #2)	4	4	4	4	4	4	4	4	4	4	4	4
ITK	4	4	4	4	4	4	4	4	4	4	4	4
LYN (ORF #1)	4	4	4	4	4	4	4	4	4	4	4	4
LYN (ORF #3)	4	4	4	4	4	4	4	4	4	4	4	4
MAPK8IP1	4	4	4	4	4	4	4	4	4	4	4	4
MIST	4	4	4	3	4	4	4	4	4	4	4	4
PIK3R1	4	4	4	4	4	3	4	4	4	4	4	4
PLCG1	4	4	4	4	4	4	4	4	4	4	4	4
PTK6	4	4	4	4	4	4	4	4	4	4	4	4
RIN2	4	4	4	4	4	4	4	4	4	4	4	4

Round Two (continued)

Kinase:	**ABL2**				**FYN**				**TNK1**			
Parent strain:	**U2**		**L40c**		**U2**		**L40c**		**U2**		**L40c**	
NLS:	-	+	-	+	-	+	-	+	-	+	-	+
ORF:												
RIN3 (ORF #1)	4	4	4	4	4	4	4	4	4	4	4	4
RIN3 (ORF #2)	4	4	4	4	4	4	4	4	4	4	4	4
RIN3 (ORF #3)	4	4	4	4	4	4	4	4	4	4	4	4
SH2D5	4	4	4	4	4	4	4	4	4	4	4	4
SH3BP2 (ORF #1)	4	4	4	4	4	4	4	4	4	4	4	4
SHC3	4	4	4	4	4	4	4	4	4	4	4	4
SHE	4	4	4	4	4	4	4	4	4	4	4	4
SLA	4	4	4	4	4	4	4	4	4	4	4	4
SOCS1	4	4	4	4	4	4	4	4	4	4	4	4
SOCS3	4	4	4	4	4	4	4	4	4	4	4	4
SRMS	4	4	4	4	4	4	4	4	4	4	4	4
STAP2	4	4	4	4	4	4	4	4	4	4	4	4
STAT3 (ORF #1)	4	4	4	4	4	4	4	4	4	4	4	2
STAT3 (ORF #2)	4	4	4	3	4	4	4	4	4	2	4	2
TBC1D4	4	4	4	4	4	4	4	4	4	4	4	4
TENC1	4	4	4	4	4	4	4	4	4	4	4	4
TNS4	4	4	4	4	4	4	4	3	4	4	4	4
VAV2	4	4	4	4	4	4	4	4	4	4	4	4
VAV3	4	4	4	4	4	4	4	4	4	4	4	4

Appendix D

Yeast two-hybrid kinase-interaction specificity

For each phosphotyrosine-dependent protein-protein interaction detected in this study the kinases that enabled growth in the yeast two-hybrid system are provided. The interactions are sorted alphabetically by the phosphotyrosine-recognizing domain-containing partner. Interactions among two phosphotyrosine-recognizing domain-containing partners are listed twice. Boldface font indicates phosphotyrosine-recognizing domain-containing genes.

Interaction		Kinases
ABL2	**CRK**	ABL2, FYN, TNK1
ABL2	**PIK3R3**	BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
ABL2	**STAT3**	ABL2, TNK1
ANKS1A	**PIK3R3**	ABL2
APBB3	PDPK1	FYN
APPL1	BATF3	ABL2
APPL1	BRWD1	ABL2, TNK1
APPL1	C22orf28	ABL2
APPL1	**CBL**	ABL2, FYN
APPL1	**CBLB**	ABL2
APPL1	**DOK2**	ABL2
APPL1	**DOK3**	ABL2
APPL1	**DOK7**	ABL2

Interaction		Kinases
APPL1	ID1	ABL2
APPL1	INO80E	ABL2
APPL1	KLF15	ABL2, TNK1
APPL1	LOC401296	ABL2
APPL1	MAGEC3	ABL2, TNK1
APPL1	**PIK3R1**	ABL2
APPL1	**SH2D2A**	ABL2
APPL1	**SOCS6**	ABL2
APPL1	ZNF829	ABL2
APPL2	GBL	ABL2, BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
APPL2	LGALS9C	FES, TNK1
APPL2	LIX1	FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
APPL2	MAPRE3	FRK, JAK2, TNK1
APPL2	MLL3	FYN
BCAR3	**IRS1**	ABL2, BLK, FER, FGR, FRK, FYN, HCK, LYN, PTK2, SRC, TNK1, YES1
BTK	**PIK3R3**	FYN
CBL	**APPL1**	ABL2, FYN
CBL	**CRK**	ABL2, FYN, TNK1
CBL	**EPS8**	ABL2
CBL	**PIK3R1**	ABL2, FRK, FYN
CBL	**PIK3R2**	ABL2, FYN
CBL	**PIK3R3**	ABL2, BLK, FER, FGR, FRK, FYN, HCK, SRC, SYK, YES1
CBL	**STAT3**	ABL2, TNK1
CBL	**STAT5A**	BMX
CBLB	**APPL1**	ABL2
CBLB	**CRK**	ABL2, FYN, TNK1
CBLB	**CRKL**	ABL2, FYN, TNK1
CBLC	**CRK**	ABL2
CCM2	**DOK4**	FYN

Interaction		Kinases
CCM2	TWIST2	BMX, FRK, FYN
CCM2	VSTM2L	TNK1
CRK	**ABL2**	ABL2, FYN, TNK1
CRK	BEX5	ABL2
CRK	C14orf65	ABL2
CRK	C22orf28	ABL2
CRK	C3orf36	ABL2, TNK1
CRK	**CBL**	ABL2, FYN, TNK1
CRK	**CBLB**	ABL2, FYN, TNK1
CRK	**CBLC**	ABL2
CRK	CHTF18	ABL2, FYN, TNK1
CRK	CNDP2	ABL2, TNK1
CRK	**DOK2**	ABL2, FYN
CRK	**DOK3**	ABL2, FYN, TNK1
CRK	**DOK7**	ABL2, FYN, TNK1
CRK	ELK1	ABL2
CRK	EPYC	ABL2
CRK	EYA3	ABL2, FYN
CRK	FAM59A	ABL2, FYN, TNK1
CRK	FSTL1	ABL2, BMX, FRK, JAK2, PTK2, SYK, TNK1
CRK	GABPB2	ABL2, FYN, TNK1
CRK	**HSH2D**	ABL2, FYN
CRK	KCTD13	ABL2, FYN, TNK1
CRK	KLHL20	ABL2, FYN
CRK	LASP1	ABL2, FYN, TNK1
CRK	LOC401296	ABL2, FYN, TNK1
CRK	MAGEC3	ABL2, FYN, TNK1
CRK	MPG	BMX, FES, FRK, FYN, JAK2, PTK2, TNK1
CRK	MYLIP	ABL2
CRK	MYOZ2	ABL2, FYN
CRK	NUFIP2	ABL2

Interaction		Kinases
CRK	OFCC1	ABL2, FYN, TNK1
CRK	PAFAH1B2	ABL2
CRK	PHC2	ABL2, FYN, TNK1
CRK	**PIK3R1**	ABL2, FYN, TNK1
CRK	**PIK3R2**	ABL2, TNK1
CRK	**PIK3R3**	ABL2, FYN, TNK1
CRK	PRKACA	ABL2
CRK	PRRG2	ABL2, FES, JAK2
CRK	PTK2	FYN
CRK	RAB2B	ABL2, TNK1
CRK	SEMA4D	ABL2
CRK	**SH2D2A**	ABL2, FYN, TNK1
CRK	**SOCS1**	ABL2
CRK	**SOCS6**	ABL2
CRK	**STAT4**	ABL2, BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
CRK	TCAP	ABL2, FYN, TNK1
CRK	TM4SF19	ABL2, TNK1
CRK	TUG1	ABL2, FYN, TNK1
CRK	VAC14	FYN
CRK	XAGE1XX	ABL2, TNK1
CRKL	**CBLB**	ABL2, FYN, TNK1
CRKL	**DOK1**	FYN
CRKL	**DOK2**	ABL2, FYN
CRKL	KIAA0317	ABL2, FYN, TNK1
CRKL	**SOCS1**	ABL2, FYN, TNK1
CRKL	TMEM168	ABL2
CSK	PDPK1	FYN
DAPP1	GSTCD	ABL2, TNK1
DAPP1	MYLIP	ABL2, TNK1
DAPP1	TNFAIP1	ABL2, FYN

Interaction		Kinases
DAPP1	WDR20	ABL2, BLK, FER, FGR, FRK, FYN, HCK, LYN, SRC, SYK, YES1
DAPP1	ZNHIT1	TNK1
DOK1	**CRKL**	FYN
DOK1	**DOK3**	ABL2, FYN
DOK2	**APPL1**	ABL2
DOK2	**CRK**	ABL2, FYN
DOK2	**CRKL**	ABL2, FYN
DOK2	**DOK3**	ABL2
DOK2	**LCK**	ABL2, FYN
DOK2	**STAT3**	ABL2, TNK1
DOK3	**APPL1**	ABL2
DOK3	**CRK**	ABL2, FYN, TNK1
DOK3	**DOK1**	ABL2, FYN
DOK3	**DOK2**	ABL2
DOK3	**DOK4**	ABL2, BLK, FER, FGR, FYN, HCK, SYK, YES1
DOK3	**DOK5**	FYN
DOK3	**SHE**	ABL2
DOK3	**STAT3**	ABL2, FYN, TNK1
DOK4	**CCM2**	FYN
DOK4	**DOK3**	ABL2, BLK, FER, FGR, FYN, HCK, SYK, YES1
DOK4	PTK2	BLK, FGR, FYN, HCK, SRC, YES1
DOK4	**STAT2**	FYN
DOK5	**DOK3**	FYN
DOK5	SCOC	ABL2, FES, PTK2
DOK7	**APPL1**	ABL2
DOK7	**CRK**	ABL2, FYN, TNK1
DOK7	MPZL1	ABL2, FYN, TNK1
EPS8	**CBL**	ABL2
EPS8	XAGE2XX	FYN
FER	KCTD4	ABL2
FRS3	**MATK**	FYN

Interaction		Kinases
FRS3	**NUMB**	FYN
FRS3	**PIK3R3**	FYN
FRS3	VASP	FYN
FYN	FYB	FYN
GRB10	BAIAP2	ABL2, BMX, FES, FRK, PTK2, SYK, TNK1
GRB10	RCAN3	ABL2
GRB14	MST4	ABL2
GRB2	ACAP1	ABL2, FYN, SYK
GRB2	C10orf81	ABL2, BLK, FER, FGR, FYN, HCK, SRC, SYK, TNK1, YES1
GRB2	C1orf135	ABL2, BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
GRB2	C6orf125	ABL2, BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
GRB2	CRBN	ABL2, BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
GRB2	DBN1	ABL2
GRB2	E2F2	ABL2, FYN
GRB2	LMX1A	ABL2, FER, FGR, FYN, YES1
GRB2	MYOZ1	ABL2
GRB2	PACRGL	ABL2, BMX, FES, FRK, FYN, JAK2, TNK1
GRB2	PARD6A	ABL2, FER, FGR, FYN, HCK, SYK, YES1
GRB2	PCDHB1	ABL2, BLK, FER, FGR, FYN, HCK, SYK, YES1
GRB2	PCDHB5	ABL2, BMX, FES, FRK, FYN, JAK2
GRB2	PPARA	FER, FGR, FRK, FYN
GRB2	RB1	ABL2, FER, FGR, FYN, HCK, YES1
GRB2	**SOCS4**	FYN
GRB2	TMEM128	ABL2, BMX, FYN, SYK
GRB2	TSPAN2	ABL2, BLK, FER, FGR, FYN, HCK, PTK2, TNK1, YES1
GRB2	VCP	ABL2, BLK, BMX, FER, FES, FGR, FRK, FYN, HCK, JAK2, PTK2, SRC, SYK, TNK1, YES1
GRB2	WDFY3	ABL2, BMX, FES, FYN, JAK2, PTK2, SYK, TNK1

Interaction		Kinases
GRB2	XAGE1XX	ABL2, FES, FRK, FYN, JAK2, PTK2, SYK
GRB7	OLIG1	ABL2, FER, FES
GRB7	PTK2	ABL2, PTK2
GRB7	ZNHIT1	FRK, SYK
HCK	NGEF	ABL2, FRK, FYN, JAK2, PTK2, SYK, TNK1
HCK	**PIK3R3**	FYN
HSH2D	**CRK**	ABL2, FYN
HSH2D	LDHAL6B	ABL2, FYN
HSH2D	LECT1	ABL2
HSH2D	**PIK3R3**	BLK, FER, FGR, FYN, YES1
HSH2D	PINK1	ABL2
HSH2D	PPAPDC2	ABL2
HSH2D	RBP7	ABL2
HSH2D	TSC1	ABL2
IRS1	**BCAR3**	ABL2, BLK, FER, FGR, FRK, FYN, HCK, LYN, PTK2, SRC, TNK1, YES1
IRS1	**NUMB**	ABL2, BLK, FER, FGR, FRK, FYN, HCK, PTK2, SRC, TNK1, YES1
IRS1	**PIK3R1**	ABL2, FRK, FYN, PTK2, TNK1
IRS1	**PIK3R3**	ABL2, BLK, FER, FGR, FRK, FYN, HCK, PTK2, SRC, TNK1, YES1
ITGB1BP1	PGRMC1	FER, FGR, FYN
JAK3	LNX1	ABL2, SYK
LCK	**DOK2**	ABL2, FYN
LCK	**STAT3**	ABL2
LCP2	FYB	BLK, FER, FGR, FYN, HCK, LCK, LYN, SRC, SYK, YES1
MAPK8IP2	ANGPT1	FRK, FYN, TNK1
MAPK8IP2	C22orf39	FRK, PTK2
MAPK8IP2	C8orf33	FRK
MAPK8IP2	FLJ21463	ABL2, BMX, FES, FRK, PTK2, SYK, TNK1
MAPK8IP2	LOC492311	PTK2
MAPK8IP2	MIA3	FES, FRK, FYN, PTK2

Interaction		Kinases
MAPK8IP2	TMEM128	ABL2
MAPK8IP2	UBD	ABL2, JAK2, PTK2, TNK1
MATK	**FRS3**	FYN
MIST	ARAP3	ABL2
MIST	C6orf146	ABL2, TNK1
MIST	COQ9	ABL2
MIST	CYB561D2	ABL2
MIST	GLIS3	ABL2, FYN, TNK1
MIST	NME4	ABL2, FYN, TNK1
MIST	**PIK3R1**	ABL2, FYN, TNK1
MIST	**PIK3R3**	ABL2, FYN, TNK1
MIST	PPP1R7	ABL2
MIST	PTK7	ABL2
MIST	RPAIN	ABL2
MIST	SFMBT1	ABL2, FYN, TNK1
MIST	**SOCS6**	ABL2, FYN, TNK1
NCK1	OLIG1	ABL2, FER
NCK1	**PIK3R3**	ABL2, BLK, FYN, SYK
NCK1	**SH2D1A**	BLK, FER, FGR, FYN, HCK, LYN, SRC
NCK2	FYB	FYN, YES1
NUMB	**FRS3**	FYN
NUMB	**IRS1**	ABL2, BLK, FER, FGR, FRK, FYN, HCK, PTK2, SRC, TNK1, YES1
NUMB	OLIG1	ABL2
PIK3R1	**APPL1**	ABL2
PIK3R1	**CBL**	ABL2, FRK, FYN
PIK3R1	**CRK**	ABL2, FYN, TNK1
PIK3R1	**IRS1**	ABL2, FRK, FYN, PTK2, TNK1
PIK3R1	**MIST**	ABL2, FYN, TNK1
PIK3R1	**SH2D2A**	ABL2, FYN
PIK3R1	**SOCS1**	ABL2, FYN, TNK1
PIK3R1	**STAT3**	ABL2

Interaction		Kinases
PIK3R1	**TNS4**	ABL2, FYN
PIK3R2	**CBL**	ABL2, FYN
PIK3R2	**CRK**	ABL2, TNK1
PIK3R2	**STAT3**	ABL2
PIK3R3	**ABL2**	BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
PIK3R3	**ANKS1A**	ABL2
PIK3R3	ARL6IP4	ABL2, FER, FGR, FYN, HCK, LYN, YES1
PIK3R3	**BTK**	FYN
PIK3R3	C10orf81	ABL2, BLK, BMX, FER, FES, FGR, FRK, FYN, HCK, SRC, SYK, TNK1, YES1
PIK3R3	**CBL**	ABL2, BLK, FER, FGR, FRK, FYN, HCK, SRC, SYK, YES1
PIK3R3	CCDC14	ABL2, BLK, FER, FES, FGR, FYN, HCK, SYK, YES1
PIK3R3	**CRK**	ABL2, FYN, TNK1
PIK3R3	CRYBA2	ABL2
PIK3R3	E2F6	ABL2, FYN
PIK3R3	EVI1	ABL2, BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
PIK3R3	**FRS3**	FYN
PIK3R3	**HCK**	FYN
PIK3R3	**HSH2D**	BLK, FER, FGR, FYN, YES1
PIK3R3	IQUB	ABL2, BMX, FES, FRK, FYN, JAK2, PTK2, TNK1
PIK3R3	**IRS1**	ABL2, BLK, FER, FGR, FRK, FYN, HCK, PTK2, SRC, TNK1, YES1
PIK3R3	**MIST**	ABL2, FYN, TNK1
PIK3R3	**NCK1**	ABL2, BLK, FYN, SYK
PIK3R3	OTUD7B	ABL2, FYN
PIK3R3	PARD6A	ABL2, FYN
PIK3R3	PCDHB5	ABL2, BMX, FES, FYN, JAK2, SYK, TNK1
PIK3R3	PELI3	ABL2, BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1

Interaction		Kinases
PIK3R3	PELO	ABL2, BLK, BMX, FES, FGR, FRK, FYN, HCK, PTK2, TNK1, YES1
PIK3R3	PLB1	ABL2, BMX, FES, FYN, JAK2, SYK, TNK1
PIK3R3	PPARA	ABL2, FYN
PIK3R3	RBP7	ABL2, BMX, FYN
PIK3R3	**SH2D2A**	ABL2, FYN
PIK3R3	**SOCS4**	ABL2
PIK3R3	SPSB2	FES
PIK3R3	**SRC**	FYN
PIK3R3	**STAT3**	ABL2, TNK1
PIK3R3	USP2	ABL2, FES, FYN, JAK2, SYK, TNK1
PIK3R3	VCP	ABL2, FYN
PIK3R3	WBSCR27	ABL2
PIK3R3	ZNF281	ABL2, FYN, TNK1
PIK3R3	ZNF767	ABL2
PLCG2	ELK1	BMX
PLCG2	EPYC	FRK, FYN, PTK2, SYK, TNK1
PLCG2	HSPD1	BMX
PLCG2	LHX8	BMX, FRK, JAK2, PTK2, SYK, TNK1
PLCG2	PTTG1	FES, PTK2
PLCG2	TP53	FRK, FYN, SYK
PTPN11	LNX1	ABL2, FES, FRK, PTK2
PTPN6	OLIG1	ABL2
PTPN6	PTK2	ABL2, FES
RABGAP1	**STAT3**	ABL2, FYN, TNK1
RABGAP1L	ACAP1	ABL2
RABGAP1L	ARFGAP1	BMX
RABGAP1L	C6orf125	ABL2, FRK, JAK2
RABGAP1L	C9orf43	ABL2, FES
RABGAP1L	DBN1	ABL2
RABGAP1L	EVI1	ABL2, BMX, FES, FRK, SYK
RABGAP1L	RB1	ABL2, FER, FGR, FRK, FYN

Interaction		Kinases
RABGAP1L	TSPAN2	ABL2, FES
RABGAP1L	WDFY3	ABL2, BMX
RASA1	OLIG1	FER, TNK1
SH2B1	CSAD	FRK, TNK1
SH2D1A	ATF3	ABL2, BMX, JAK2, PTK2, SYK, TNK1
SH2D1A	**NCK1**	BLK, FER, FGR, FYN, HCK, LYN, SRC
SH2D1A	PTK2	ABL2, FYN, TNK1
SH2D1A	ZNHIT1	ABL2, BLK, FER, FGR, FYN, HCK
SH2D1B	HSPA1A	ABL2, BLK, BMX, FGR, FYN, HCK, SRC, TNK1, YES1
SH2D1B	**STAT4**	ABL2, BLK, FYN
SH2D2A	**APPL1**	ABL2
SH2D2A	ASB3	ABL2
SH2D2A	**CRK**	ABL2, FYN, TNK1
SH2D2A	FAM46A	ABL2
SH2D2A	FAM46B	ABL2, FES, FRK, FYN, JAK2, TNK1
SH2D2A	KLRAQ1	ABL2
SH2D2A	LNX1	ABL2
SH2D2A	LRRFIP1	ABL2, CSK, FER, FGR, FYN, JAK3, TNK1
SH2D2A	**PIK3R1**	ABL2, FYN
SH2D2A	**PIK3R3**	ABL2, FYN
SH2D2A	PLEKHB1	ABL2, FRK, FYN, JAK2, SYK
SH2D2A	PTK2	BLK, FGR, FYN, HCK, SRC, SYK, YES1
SH2D2A	RAD54B	ABL2, TNK1
SH2D2A	**SOCS3**	ABL2, FYN
SH2D2A	SPTB	ABL2, BLK, FER, FGR, TNK1, YES1
SH2D2A	**STAT3**	ABL2, FYN
SH2D2A	TMEM148	ABL2
SH2D2A	TSC1	ABL2, BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
SH2D2A	WDR40B	ABL2
SH2D2A	ZHX3	ABL2

Interaction		Kinases
SH2D2A	ZSCAN1	ABL2, FYN
SH3BP2	ARHGAP10	FYN
SH3BP2	**STAT3**	ABL2, TNK1
SHD	CAPN10	ABL2, BMX, FES, FRK, JAK2, PTK2, SYK, TNK1
SHD	PSG11	ABL2, BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
SHE	**DOK3**	ABL2
SHE	TRIT1	ABL2
SOCS1	**CRK**	ABL2
SOCS1	**CRKL**	ABL2, FYN, TNK1
SOCS1	**PIK3R1**	ABL2, FYN, TNK1
SOCS3	**SH2D2A**	ABL2, FYN
SOCS3	TRDN	FYN
SOCS4	**GRB2**	FYN
SOCS4	LDHAL6B	ABL2, PTK2
SOCS4	LECT1	ABL2
SOCS4	MPP5	ABL2, FES, TNK1
SOCS4	**PIK3R3**	ABL2
SOCS4	PINK1	ABL2, PTK2
SOCS4	PPAPDC2	ABL2, PTK2
SOCS4	RBP7	ABL2, PTK2
SOCS4	TNK2	BMX, FRK, PTK2, TNK1
SOCS6	**APPL1**	ABL2
SOCS6	**CRK**	ABL2
SOCS6	LNX1	BMX, FES, FRK, FYN, PTK2, TNK1
SOCS6	**MIST**	ABL2, FYN, TNK1
SRC	**PIK3R3**	FYN
STAT1	SPTB	ABL2, BMX, FES, FYN, JAK2
STAT2	**DOK4**	FYN
STAT3	**ABL2**	ABL2, TNK1
STAT3	BRWD1	ABL2, FYN

Interaction		Kinases
STAT3	C10orf18	ABL2, FYN, TNK1
STAT3	CA8	ABL2
STAT3	**CBL**	ABL2, TNK1
STAT3	CCDC87	ABL2, TNK1
STAT3	**DOK2**	ABL2, TNK1
STAT3	**DOK3**	ABL2, FYN, TNK1
STAT3	DTNA	ABL2, FYN, TNK1
STAT3	FAM117B	ABL2, FYN, TNK1
STAT3	GSTCD	ABL2, FYN, TNK1
STAT3	HESX1	ABL2, FYN, TNK1
STAT3	KRTAP10-7	ABL2, FYN, TNK1
STAT3	LASP1	ABL2, FYN, TNK1
STAT3	**LCK**	ABL2
STAT3	LOC401296	ABL2, FYN, TNK1
STAT3	MAPKAPK2	ABL2
STAT3	MPZL1	ABL2, FYN, TNK1
STAT3	NUFIP2	ABL2, FYN
STAT3	PAQR7	ABL2, TNK1
STAT3	**PIK3R1**	ABL2
STAT3	**PIK3R2**	ABL2
STAT3	**PIK3R3**	ABL2, TNK1
STAT3	PPARD	ABL2, TNK1
STAT3	**RABGAP1**	ABL2, FYN, TNK1
STAT3	**SH2D2A**	ABL2, FYN
STAT3	**SH3BP2**	ABL2, TNK1
STAT3	TM4SF19	ABL2, FYN, TNK1
STAT3	TWIST1	ABL2, FYN
STAT3	VPS39	ABL2, TNK1
STAT3	WDFY3	ABL2, FYN, TNK1
STAT3	ZNF281	ABL2, FYN, TNK1
STAT3	ZNF829	ABL2

Interaction		Kinases
STAT4	**CRK**	ABL2, BMX, FES, FRK, FYN, JAK2, PTK2, SYK, TNK1
STAT4	**SH2D1B**	ABL2, BLK, FYN
STAT5A	**CBL**	BMX
STAT5A	OLIG1	ABL2, BMX, JAK2, PTK2, SYK, TNK1
SYK	FAM46A	FES
TNS4	GSTCD	ABL2, FYN, TNK1
TNS4	NME4	ABL2, FYN
TNS4	P4HA2	ABL2, FYN, TNK1
TNS4	**PIK3R1**	ABL2, FYN
TNS4	RPAIN	ABL2
YES1	BECN1	ABL2, BMX, FER, FES, FGR, HCK, JAK2, PTK2, SYK, TNK1, YES1
YES1	**OLIG1**	ABL2, BLK, FER, FES, FGR, HCK, PTK2, YES1

Appendix E

Mutant SH2 CoIP results

Co-immunoprecipitation assays employing binding-deficient SH2 domains demonstrate phosphotyrosine-dependency regardless of high background tyrosine kinase activity. Thirty interactions validated in co-immunoprecipitation assays were selected and repeated with constructs containing mutated SH2 domains. Interaction phosphotyrosine-dependency was considered validated if the luciferase signal was reduced by a factor of at least two and by at least two standard deviations compared to wildtype signal, assuming a multiplicative error model. The position of each ORF ("Protein A" or "Luciferase" for pPAReni-DM or pFireV5-DM, respectively), the binding signal strength ("Fold Binding") and its statistical reliability in terms of standard deviations ("Z-Score") are provided for each pair of ORFs.

Protein A	Luciferase	Fold Binding	Z-Score
PIK3R3 (ORF #2 R90LR383L)	IRS1	$0.03 \,^{\times}\!\div 1.16$	-23.6
PIK3R1 (R88KR379K)	IRS1	$0.07 \,^{\times}\!\div 1.22$	-13.6
GRB2 (ORF #2 R86K)	RB1	$0.09 \,^{\times}\!\div 1.14$	-18.3
PIK3R3 (ORF #2 R90LR383L)	WDFY3	$0.12 \,^{\times}\!\div 1.88$	-3.4
PIK3R3 (ORF #2 R90LR383L)	CRYBA2	$0.15 \,^{\times}\!\div 1.21$	-9.8
PIK3R3 (ORF #2 R90LR383L)	PELI3	$0.18 \,^{\times}\!\div 1.04$	-39.3
PIK3R3 (ORF #2 R90LR383L)	PARD6A	$0.22 \,^{\times}\!\div 1.16$	-10.5
FRS3 (ORF #2)	SOCS4 (R311L)	$0.23 \,^{\times}\!\div 1.15$	-10.3
PIK3R3 (ORF #2 R90LR383L)	SH2D2A	$0.24 \,^{\times}\!\div 1.25$	-6.4
PIK3R3 (ORF #2 R90LR383L)	SPSB2	$0.27 \,^{\times}\!\div 1.17$	-8.6

Protein A — Luciferase	Fold Binding	Z-Score
GRB2 (ORF #2 R86K) — TSPAN2	$0.27^{\times}_{\div}1.35$	-4.4
C10orf81 — PIK3R3 (ORF #2 R90LR383L)	$0.29^{\times}_{\div}1.27$	-5.1
GRB2 (ORF #2 R86K) — VCP (ORF #1)	$0.32^{\times}_{\div}1.17$	-7.1
PIK3R3 (ORF #2 R90LR383L) — HCK	$0.35^{\times}_{\div}1.48$	-2.7
GRB2 (ORF #2 R86K) — TMEM128	$0.35^{\times}_{\div}1.19$	-5.9
PIK3R3 (ORF #2 R90LR383L) — WBSCR27	$0.35^{\times}_{\div}1.34$	-3.6
BCAR3 (R177L) — IRS1	$0.44^{\times}_{\div}1.33$	-2.8
PIK3R3 (ORF #2 R90LR383L) — OTUD7B	$0.46^{\times}_{\div}1.39$	-2.4
PIK3R1 (R88KR379K) — SH2D2A	$0.48^{\times}_{\div}1.23$	-3.5
PIK3R3 (ORF #2 R90LR383L) — ARL6IP4	$0.48^{\times}_{\div}1.38$	-2.3
ZHX3 — SH2D2A (R120K)	$0.59^{\times}_{\div}1.19$	-3
PIK3R3 (ORF #2 R90LR383L) — PPARA	$0.6^{\times}_{\div}1.56$	-1.1
CRK (ORF #3) — SH2D2A (R120K)	$0.65^{\times}_{\div}1.17$	-2.6
GRB2 (ORF #2 R86K) — SOCS4	$0.66^{\times}_{\div}1.42$	-1.2
PINK1 — SOCS4 (R311K)	$0.88^{\times}_{\div}1.16$	-0.9
SH2D2A (R120K) — FAM46A	$0.89^{\times}_{\div}1.22$	-0.6
GRB2 (ORF #2 R86K) — DBN1	$0.89^{\times}_{\div}1.41$	-0.3
GRB2 (ORF #2 R86K) — PPARA	$0.93^{\times}_{\div}1.4$	-0.2
SH2D2A (R120K) — FAM46B	$0.94^{\times}_{\div}1.25$	-0.3
SH2D2A (R120K) — ASB3	$0.95^{\times}_{\div}1.29$	-0.2
PLEKHB1 — SH2D2A (R120K)	$0.95^{\times}_{\div}1.17$	-0.3
RBP7 — SOCS4 (R311K)	$0.96^{\times}_{\div}1.06$	-0.6
GRB2 (ORF #2 R86K) — LMX1A	$0.96^{\times}_{\div}1.52$	-0.1
SH2D2A (R120K) — PTK2 (ORF #1 WT)	$1.04^{\times}_{\div}1.19$	0.2
GRB2 (ORF #2 R86K) — C10orf81	$1.1^{\times}_{\div}1.11$	0.9
RAD54B — SH2D2A (R120K)	$1.15^{\times}_{\div}1.28$	0.6
LRRFIP1 — SH2D2A (R120K)	$1.19^{\times}_{\div}1.25$	0.8
SH2D2A (R120K) — WDR40B	$1.19^{\times}_{\div}1.22$	0.9
TSC1 (ORF #2) — SH2D2A (R120K)	$1.22^{\times}_{\div}2.49$	0.2
PIK3R3 (ORF #2 R90LR383L) — SOCS4	$1.47^{\times}_{\div}1.12$	3.4

Appendix F

Protein Complementation assay results

Results of the protein complementation assay experiment including observed subcellular localization and exact configuration of ORFs and fragment fusions. For each combination of ORF and fragment fusion plasmid not involved in an experiment with negative outcome, a control experiment is provided. pos = positive; neg = negative; ctrl = control; C = carboxy-terminal fusion; N = amino-terminal fusion.

Result	Localization	F1 fusion		F2 fusion	
pos	nucleus	GRB2 (ORF #2)	C	ARL6IP4	N
pos	cytoplasm	GRB2 (ORF #2)	C	C10orf81	C
pos	cytoplasm	GRB2 (ORF #2)	C	PARD6A	C
pos	nucleus	GRB2 (ORF #2)	C	RB1	N
pos	nucleus	GRB2 (ORF #2)	C	RB1	C
pos	membrane	GRB2 (ORF #2)	C	TSPAN2	C
pos	nucleus	PIK3R3 (ORF #1)	C	ARL6IP4	C
pos	cytoplasm	PIK3R3 (ORF #1)	C	C10orf81	C
pos	cytoplasm	PIK3R3 (ORF #1)	C	C10orf81	C
pos	cytoplasm	PIK3R3 (ORF #1)	C	TSPAN2	C
pos	nucleus	PIK3R3 (ORF #4)	C	RB1	N
pos	nucleus	PIK3R3 (ORF #4)	C	SOCS4	C

Result	Localization	F1 fusion		F2 fusion	
pos	cytoplasm	RBP7	N	PIK3R3 (ORF #4)	C
pos	nucleus+cytoplasm	SOCS4	N	RBP7	N
neg	perinuclear cloud	GRB2 (ORF #2)	C	OLIG1	N
neg	perinuclear cloud	GRB2 (ORF #2)	C	OLIG1	C
neg	no signal	GRB2 (ORF #2)	C	SOCS4	C
neg	perinuclear cloud	GRB2 (ORF #2)	C	TMEM128	N
neg	perinuclear cloud	GRB2 (ORF #2)	C	TMEM128	C
neg	perinuclear cloud	GRB2 (ORF #2)	C	WBSCR27	N
neg	perinuclear cloud	GRB2 (ORF #2)	C	WBSCR27	C
neg	no signal	PIK3R1	N	SH2D2A	N
neg	no signal	PIK3R3 (ORF #1)	C	SH2D2A	N
neg	no signal	PTK2 (ORF #1 WT)	C	SH2D2A	N
ctrl	no signal	–	-	PARD6A	C
ctrl	no signal	–	-	TSPAN2	C
ctrl	no signal	ARL6IP4	N	PIK3R3 (ORF #4)	C
ctrl	no signal	GRB2 (ORF #2)	C	ARL6IP4	C
ctrl	no signal	GRB2 (SH2)	N	ARL6IP4	N
ctrl	no signal	GRB2 (SH2)	N	C10orf81	C
ctrl	no signal	GRB2 (SH2)	N	RB1	N
ctrl	no signal	PIK3R3 (ORF #4)	C	RB1	C
ctrl	no signal	PIK3R3 (ORF #4)	C	RB1	C
ctrl	no signal	RBP7	N	PIK3R3 (ORF #1 R90LR383L)	C
ctrl	no signal	SOCS4	N	RBP7	C
ctrl	no signal	SOCS4	C	RBP7	N

Appendix G

Known interactions recovered

For each interaction detected in this study that has been described before, the phosphotyrosine-dependency in this study (pY-dependent this study) and, where available, in the previous description (pY-dependent literature) are indicated. For greater clarity, each (non-self-)interaction is listed twice.

Interaction	pY-dependent this study	pY-dependent literature
ABL2 — CRK	yes	no
AKT1 — APPL1	no	–
APPL1 — AKT1	no	–
BAIAP2 — EPS8	no	no
BMX — STAT3	no	yes
CBL — CRK	yes	yes
CBL — CRKL	no	both
CBL — FYN	no	no
CBL — GRB2	no	no
CBL — PIK3R1	yes	yes
CBL — PIK3R2	yes	–
CBLB — CRK	yes	–
CBLB — CRKL	yes	yes
CBLB — FYN	no	–
CBLC — CRK	yes	no
CCM2 — KRIT1	no	–
CRK — ABL2	yes	no
CRK — CBL	yes	yes

Interaction	pY-dependent	
	this study	literature
CRK — CBLB	yes	–
CRK — CBLC	yes	no
CRK — PIK3R1	yes	–
CRK — PIK3R2	yes	–
CRK — PIK3R3	yes	–
CRK — PTK2	yes	yes
CRK — RIN3	no	no
CRK — SOCS1	yes	no
CRKL — CBL	no	both
CRKL — CBLB	yes	yes
CRKL — DOK1	yes	yes
DOK1 — CRKL	yes	yes
DOK2 — LCK	yes	–
EPS8 — BAIAP2	no	no
FGF12 — MAPK8IP2	no	–
FGF13 — MAPK8IP2	no	–
FYN — CBL	no	no
FYN — CBLB	no	–
GRB2 — CBL	no	no
GRB7 — PTK2	yes	yes
HSH2D — TNK2	no	–
IRS1 — PIK3R1	yes	yes
IRS1 — PIK3R3	yes	–
ITGB1BP1 — KRIT1	no	–
KRIT1 — CCM2	no	–
KRIT1 — ITGB1BP1	no	–
LCK — DOK2	yes	–
LNX2 — NUMB	no	yes
LYN — PTK2	no	–
MAPK8IP2 — FGF12	no	–
MAPK8IP2 — FGF13	no	–

Interaction	pY-dependent	
	this study	literature
NUMB — LNX2	no	yes
PIK3CA — PIK3R3	no	no
PIK3R1 — CBL	yes	yes
PIK3R1 — CRK	yes	–
PIK3R1 — IRS1	yes	yes
PIK3R1 — SOCS1	yes	no
PIK3R2 — CBL	yes	–
PIK3R2 — CRK	yes	–
PIK3R3 — CRK	yes	–
PIK3R3 — IRS1	yes	–
PIK3R3 — PIK3CA	no	no
PTK2 — CRK	yes	yes
PTK2 — GRB7	yes	yes
PTK2 — LYN	no	–
PTK2 — SRC	no	yes
RIN3 — CRK	no	no
SOCS1 — CRK	yes	no
SOCS1 — PIK3R1	yes	no
SRC — PTK2	no	yes
STAT3 — BMX	no	yes
TNK2 — HSH2D	no	–

Appendix H

Literature interactions among successful bait genes

For each interaction among successful bait genes recovered from the ConsensusPathDB meta-database, the PubMed identifiers (PMIDs) of the examined publications and the conclusion reached (validated) are listed.
Note that, due to the points of interests in this study, for a validated interaction, further publications were examined only if relevant for another interaction.

Interaction	validated	PMID(s)
ABL2 — CBL	no	19380743
ABL2 — CBLB	no	19380743
ABL2 — CRK	yes	8194526, 17474147
ABL2 — FYN	yes	17474147
ABL2 — GRB2	yes	17474147
ABL2 — HCK	yes	12748290
ABL2 — PIK3R2	no	19380743
ABL2 — PIK3R3	no	19380743
ABL2 — RIN1	yes	9144171
ABL2 — SRC	yes	17474147
APBB3 — APPL2	yes	9461550
APPL1 — APPL2	yes	16189514
APPL2 — APBB3	yes	9461550

Interaction	validated	PMID(s)
APPL2 — APPL1	yes	16189514
BTK — CBL	yes	7629518
BTK — CBLB	yes	12093870
BTK — DAPP1	yes	11524430
BTK — FYN	yes	8058772
BTK — GRB2	yes	19380743, 20936779
BTK — HCK	yes	8058772
BTK — LYN	yes	8058772
BTK — PIK3R2	no	19380743
BTK — PLCG2	yes	10981967, 12093870
BTK — PTPN11	no	19380743
BTK — STAT3	yes	17367410
BTK — STAT5A	yes	11413148
BTK — SYK	no	8629002, 8630736, 11226282, 11598012, 12522270
CBL — ABL2	no	19380743
CBL — BTK	yes	7629518
CBL — CBLB	no	15657067, 19380743, 21706016
CBL — CRK	yes	8524328, 8621483, 8626543, 9178909, 9461587, 19380743
CBL — CRKL	yes	8621483, 8626543, 8662998, 9092574, 9178909, 9344843, 9461587, 11133830, 15556646
CBL — CSK	yes	19380743
CBL — FRS2	yes	11997436
CBL — FYN	yes	7642581, 8621719, 9525940, 9535867, 9890970, 15190072, 15556646
CBL — GRB2	yes	7642581, 8621719, 8662998, 9178909, 9461587, 11133830, 11964172, 11997436, 19380743, 20936779
CBL — HCK	yes	9890970, 10092522, 10799548, 11896602
CBL — IRS1	no	21706016

Interaction	validated	PMID(s)
CBL — LYN	yes	9160881, 9890970, 15190072
CBL — PIK3R2	yes	11133830
CBL — PIK3R3	no	19380743, 21706016
CBL — PTPN11	yes	11157475
CBL — PTPN6	yes	12176909
CBL — SH3BP2	yes	9846481
CBL — SRC	yes	9525940, 11994282
CBL — STAT3	no	15657067
CBL — STAT5A	no	12193575
CBL — SYK	yes	8621719, 9535867, 15556646
CBL — YES1	yes	9525940
CBLB — ABL2	no	19380743
CBLB — BTK	yes	12093870
CBLB — CBL	no	15657067, 19380743, 21706016
CBLB — CRK	yes	19380743
CBLB — CRKL	yes	10022120, 12697763
CBLB — CSK	no	21706016
CBLB — FRS2	no	21706016
CBLB — FYN	yes	16503409
CBLB — GRB2	yes	10022120, 12577067, 20936779
CBLB — HCK	no	12029088
CBLB — IRS1	yes	16734387
CBLB — PIK3R2	no	19380743, 21706016
CBLB — PIK3R3	no	19380743, 21706016
CBLB — PLCG2	yes	12093870
CBLB — PTPN11	no	12577067, 21706016
CBLB — SH3BP2	no	17306257
CBLB — STAT3	no	15657067
CBLB — SYK	yes	10022120
CBLB — YES1	no	15657067

Interaction	validated	PMID(s)
CRK — ABL2	yes	8194526, 17474147
CRK — CBL	yes	8524328, 8621483, 8626543, 9178909, 9461587, 19380743
CRK — CBLB	yes	19380743
CRK — CRKL	no	16713569
CRK — CSK	no	19380743
CRK — DOK4	yes	12730241
CRK — FRS2	yes	10092678
CRK — FYN	no	9480911, 9642287, 11956190, 12198159
CRK — GRB2	yes	19380743
CRK — IRS1	yes	8621590
CRK — PIK3R2	yes	19380743
CRK — PIK3R3	yes	19380743
CRK — PTPN11	yes	19380743
CRK — RIN3	yes	17474147
CRK — SOCS1	yes	10022833
CRK — SRC	yes	12615911
CRK — STAT5A	yes	11097834
CRKL — CBL	yes	8621483, 8626543, 8662998, 9092574, 9178909, 9344843, 9461587, 11133830, 15556646
CRKL — CBLB	yes	10022120, 12697763
CRKL — CRK	no	16713569
CRKL — DOK1	yes	11071635
CRKL — FYN	no	15556646
CRKL — GRB2	yes	10477741
CRKL — LYN	yes	11443118
CRKL — PIK3R2	yes	9092574
CRKL — PTPN11	yes	9344843
CRKL — STAT5A	yes	9837784
CRKL — SYK	yes	11313252

Interaction	validated	PMID(s)
CSK — CBL	yes	19380743
CSK — CBLB	no	21706016
CSK — CRK	no	19380743
CSK — DOK1	no	9221755, 10799545, 10852966, 11551902, 12522270, 15144186, 15592455, 16094384
CSK — FRS2	no	21706016
CSK — FYN	yes	7524477
CSK — GRB2	no	19380743
CSK — HCK	no	10934191, 11976726
CSK — IRS1	no	21706016
CSK — LYN	yes	1722201
CSK — PIK3R2	yes	19380743
CSK — PIK3R3	no	21706016
CSK — PTPN11	no	14665621, 21706016
CSK — SRC	yes	12387813
CSK — YES1	no	16094384
DAPP1 — BTK	yes	11524430
DAPP1 — LYN	yes	10880360, 11524430
DAPP1 — PLCG2	yes	10770799
DAPP1 — SRC	yes	10880360, 11524430
DOK1 — CRKL	yes	11071635
DOK1 — CSK	no	9221755, 10799545, 10852966, 11551902, 12522270, 15144186, 15592455, 16094384
DOK1 — DOK2	no	10822173
DOK1 — DOK7	no	20562859
DOK1 — FYN	yes	15345598
DOK1 — GRB10	no	11551902
DOK1 — HCK	yes	11071635
DOK1 — LYN	yes	11071635
DOK1 — PTPN11	yes	15546884

Interaction	validated	PMID(s)
DOK1 — PTPN6	yes	10585470
DOK1 — SH2D1A	yes	10852966
DOK1 — SRC	yes	11071635
DOK1 — YES1	yes	11071635
DOK2 — DOK1	no	10822173
DOK2 — HCK	no	10428862
DOK2 — LYN	no	10428862
DOK2 — SRC	no	10428862
DOK4 — CRK	yes	12730241
DOK4 — FYN	yes	12730241
DOK4 — SRC	yes	12730241
DOK7 — DOK1	no	20562859
EPS8 — SRC	yes	10395945
FER — IRS1	yes	11006284
FER — STAT3	yes	10878010
FRS2 — CBL	yes	11997436
FRS2 — CBLB	no	21706016
FRS2 — CRK	yes	10092678
FRS2 — CSK	no	21706016
FRS2 — GRB2	yes	9182757, 10092678, 11997436
FRS2 — IRS1	no	21706016
FRS2 — PIK3R2	no	21706016
FRS2 — PIK3R3	no	21706016
FRS2 — PTPN11	yes	10650943
FRS2 — SRC	yes	10092678
FRS3 — GRB2	no	11432792
FRS3 — PTPN11	yes	11432792
FYN — ABL2	yes	17474147
FYN — BTK	yes	8058772

Interaction	validated	PMID(s)
FYN — CBL	yes	7642581, 8621719, 9525940, 9535867, 9890970, 15190072, 15556646
FYN — CBLB	yes	16503409
FYN — CRK	no	9480911, 9642287, 11956190, 12198159
FYN — CRKL	no	15556646
FYN — CSK	yes	7524477
FYN — DOK1	yes	15345598
FYN — DOK4	yes	12730241
FYN — GRB10	yes	10871840
FYN — GRB2	yes	9344857
FYN — IRS1	yes	8631859
FYN — LYN	no	1544885, 1722201, 7545683, 8530369, 8612628, 15190072
FYN — PIK3R2	yes	1334406
FYN — PLCG2	yes	7682059, 8395016
FYN — PTPN11	yes	10212213
FYN — RIN3	yes	17474147
FYN — SH2D1A	yes	12545174
FYN — SH3BP2	yes	9846481
FYN — SOCS1	yes	10022833
FYN — SRC	no	9169421, 16966330
FYN — STAT1	yes	9804857
FYN — SYK	no	9535867, 12522270, 16713569
FYN — YES1	yes	12640114
GRAP2 — GRB2	no	11997510, 12176364, 12640133
GRB2 — ABL2	yes	17474147
GRB2 — BTK	yes	19380743, 20936779
GRB2 — CBL	yes	7642581, 8621719, 8662998, 9178909, 9461587, 11133830, 11964172, 11997436, 19380743, 20936779
GRB2 — CBLB	yes	10022120, 12577067, 20936779

Interaction	validated	PMID(s)
GRB2 — CRK	yes	19380743
GRB2 — CRKL	yes	10477741
GRB2 — CSK	no	19380743
GRB2 — FRS2	yes	9182757, 10092678, 11997436
GRB2 — FRS3	no	11432792
GRB2 — FYN	yes	9344857
GRB2 — GRAP2	no	11997510, 12176364, 12640133
GRB2 — GRB10	no	8798570, 17474147
GRB2 — GRB7	no	16189514
GRB2 — IRS1	yes	7862167
GRB2 — LYN	no	10469124, 16799092
GRB2 — PIK3R2	yes	8662998, 19380743, 20936779
GRB2 — PIK3R3	no	19380743
GRB2 — PLCG2	yes	17474147
GRB2 — PTPN11	yes	10212213, 10747947, 19380743
GRB2 — PTPN6	yes	8632004, 10747947, 11964172
GRB2 — RIN3	yes	17474147
GRB2 — SH2B1	yes	9742218, 20936779
GRB2 — SH2D1A	yes	9856458
GRB2 — SH2D4A	yes	20936779
GRB2 — SH3BP2	yes	9846481
GRB2 — SOCS1	yes	10022833
GRB2 — SRC	yes	11964172
GRB2 — STAT3	no	15657067
GRB2 — SYK	yes	10747947, 11964172
GRB2 — YES1	no	15657067
GRB7 — GRB2	no	16189514
GRB10 — DOK1	no	11551902
GRB10 — FYN	yes	10871840
GRB10 — GRB2	no	8798570, 17474147

Interaction	validated	PMID(s)
GRB10 — IRS1	no	8621530
GRB10 — SRC	yes	10871840
GRB14 — IRS1	no	9748281
HCK — ABL2	yes	12748290
HCK — BTK	yes	8058772
HCK — CBL	yes	9890970, 10092522, 10799548, 11896602
HCK — CBLB	no	12029088
HCK — CSK	no	10934191, 11976726
HCK — DOK1	yes	11071635
HCK — DOK2	no	10428862
HCK — PIK3R2	yes	12029088
HCK — PLCG2	yes	7682059
HCK — SRC	yes	11896602
HCK — STAT3	yes	12244095
IRS1 — CBL	no	21706016
IRS1 — CBLB	yes	16734387
IRS1 — CRK	yes	8621590
IRS1 — CSK	no	21706016
IRS1 — FER	yes	11006284
IRS1 — FRS2	no	21706016
IRS1 — FYN	yes	8631859
IRS1 — GRB10	no	8621530
IRS1 — GRB14	no	9748281
IRS1 — GRB2	yes	7862167
IRS1 — JAK3	yes	7499365
IRS1 — PIK3R2	yes	11120660
IRS1 — PIK3R3	yes	9415396
IRS1 — PTPN11	yes	8505282
IRS1 — PTPN6	no	18729074
IRS1 — SOCS1	yes	12228220

Interaction	validated	PMID(s)
IRS1 — SOCS3	no	12228220
JAK3 — IRS1	yes	7499365
JAK3 — PTPN6	no	8114715, 8692915, 9520455, 10574931
JAK3 — SOCS1	yes	11133764
JAK3 — SOCS3	yes	10373548
JAK3 — STAT3	no	8272872, 8626374, 9083098, 9305919, 9343414, 9566874, 9582023, 9872331, 10037026, 10446219, 10521505, 10918587, 11294897, 11335711, 11350938, 11429593, 11850821, 11940572, 12244095, 12576423, 12626508, 12763138, 14551213, 15465816, 15592455, 16189514
JAK3 — STAT5A	yes	9047382
LYN — BTK	yes	8058772
LYN — CBL	yes	9160881, 9890970, 15190072
LYN — CRKL	yes	11443118
LYN — CSK	yes	1722201
LYN — DAPP1	yes	10880360, 11524430
LYN — DOK1	yes	11071635
LYN — DOK2	no	10428862
LYN — FYN	no	1544885, 1722201, 7545683, 8530369, 8612628, 15190072
LYN — GRB2	no	10469124, 16799092
LYN — PLCG2	yes	7682059, 8395016, 10981967
LYN — PTPN6	yes	10574931
LYN — SRC	no	9169421
LYN — SYK	yes	7831290
LYN — YES1	no	1544885
PIK3R2 — ABL2	no	19380743
PIK3R2 — BTK	no	19380743
PIK3R2 — CBL	yes	11133830
PIK3R2 — CBLB	no	19380743, 21706016

Interaction	validated	PMID(s)
PIK3R2 — CRK	yes	19380743
PIK3R2 — CRKL	yes	9092574
PIK3R2 — CSK	yes	19380743
PIK3R2 — FRS2	no	21706016
PIK3R2 — FYN	yes	1334406
PIK3R2 — GRB2	yes	8662998, 19380743, 20936779
PIK3R2 — HCK	yes	12029088
PIK3R2 — IRS1	yes	11120660
PIK3R2 — PIK3R3	no	16456542, 19380743, 21706016
PIK3R2 — PTPN11	yes	19380743
PIK3R2 — SOCS1	yes	10022833
PIK3R2 — SOCS6	yes	12052866
PIK3R2 — SYK	yes	15536084
PIK3R2 — YES1	no	17620599
PIK3R3 — ABL2	no	19380743
PIK3R3 — CBL	no	19380743, 21706016
PIK3R3 — CBLB	no	19380743, 21706016
PIK3R3 — CRK	yes	19380743
PIK3R3 — CSK	no	21706016
PIK3R3 — FRS2	no	21706016
PIK3R3 — GRB2	no	19380743
PIK3R3 — IRS1	yes	9415396
PIK3R3 — PIK3R2	no	16456542, 19380743, 21706016
PIK3R3 — PTPN11	no	21706016
PLCG2 — BTK	yes	10981967, 12093870
PLCG2 — CBLB	yes	12093870
PLCG2 — DAPP1	yes	10770799
PLCG2 — FYN	yes	7682059, 8395016
PLCG2 — GRB2	yes	17474147
PLCG2 — HCK	yes	7682059

Interaction	validated	PMID(s)
PLCG2 — LYN	yes	7682059, 8395016, 10981967
PLCG2 — PTPN11	yes	12135708
PLCG2 — SH3BP2	yes	11390470
PLCG2 — SYK	yes	10469124, 10981967
PTPN11 — BTK	no	19380743
PTPN11 — CBL	yes	11157475
PTPN11 — CBLB	no	12577067, 21706016
PTPN11 — CRK	yes	19380743
PTPN11 — CRKL	yes	9344843
PTPN11 — CSK	no	14665621, 21706016
PTPN11 — DOK1	yes	15546884
PTPN11 — FRS2	yes	10650943
PTPN11 — FRS3	yes	11432792
PTPN11 — FYN	yes	10212213
PTPN11 — GRB2	yes	10212213, 10747947, 19380743
PTPN11 — IRS1	yes	8505282
PTPN11 — PIK3R2	yes	19380743
PTPN11 — PIK3R3	no	21706016
PTPN11 — PLCG2	yes	12135708
PTPN11 — PTPN6	no	8541543
PTPN11 — SH2D1A	no	11806999
PTPN11 — SOCS3	yes	10777583
PTPN11 — SRC	yes	14687660
PTPN11 — STAT1	yes	12270932
PTPN11 — STAT3	yes	11594781
PTPN11 — STAT5A	yes	10617656
PTPN6 — CBL	yes	12176909
PTPN6 — DOK1	yes	10585470
PTPN6 — GRB2	yes	8632004, 10747947, 11964172
PTPN6 — IRS1	no	18729074

Interaction	validated	PMID(s)
PTPN6 — JAK3	no	8114715, 8692915, 9520455, 10574931
PTPN6 — LYN	yes	10574931
PTPN6 — PTPN11	no	8541543
PTPN6 — SH3BP2	yes	16649996
PTPN6 — SRC	yes	9261115
PTPN6 — STAT3	no	15870198
PTPN6 — SYK	yes	10747947
RIN1 — ABL2	yes	9144171
RIN3 — CRK	yes	17474147
RIN3 — FYN	yes	17474147
RIN3 — GRB2	yes	17474147
RIN3 — SRC	yes	17474147
SH2B1 — GRB2	yes	9742218, 20936779
SH2D1A — DOK1	yes	10852966
SH2D1A — FYN	yes	12545174
SH2D1A — GRB2	yes	9856458
SH2D1A — PTPN11	no	11806999
SH2D1A — SH2D1B	no	11689425
SH2D1B — SH2D1A	no	11689425
SH2D2A — SRC	yes	15962004
SH2D4A — GRB2	yes	20936779
SH3BP2 — CBL	yes	9846481
SH3BP2 — CBLB	no	17306257
SH3BP2 — FYN	yes	9846481
SH3BP2 — GRB2	yes	9846481
SH3BP2 — PLCG2	yes	11390470
SH3BP2 — PTPN6	yes	16649996
SH3BP2 — SYK	yes	9846481
SOCS1 — CRK	yes	10022833
SOCS1 — FYN	yes	10022833

Interaction	validated	PMID(s)
SOCS1 — GRB2	yes	10022833
SOCS1 — IRS1	yes	12228220
SOCS1 — JAK3	yes	11133764
SOCS1 — PIK3R2	yes	10022833
SOCS3 — IRS1	no	12228220
SOCS3 — JAK3	yes	10373548
SOCS3 — PTPN11	yes	10777583
SOCS6 — PIK3R2	yes	12052866
SRC — ABL2	yes	17474147
SRC — CBL	yes	9525940, 11994282
SRC — CRK	yes	12615911
SRC — CSK	yes	12387813
SRC — DAPP1	yes	10880360, 11524430
SRC — DOK1	yes	11071635
SRC — DOK2	no	10428862
SRC — DOK4	yes	12730241
SRC — EPS8	yes	10395945
SRC — FRS2	yes	10092678
SRC — FYN	no	9169421, 16966330
SRC — GRB10	yes	10871840
SRC — GRB2	yes	11964172
SRC — HCK	yes	11896602
SRC — LYN	no	9169421
SRC — PTPN11	yes	14687660
SRC — PTPN6	yes	9261115
SRC — RIN3	yes	17474147
SRC — SH2D2A	yes	15962004
SRC — STAT1	yes	10358079
SRC — STAT3	yes	8657134, 12244095
SRC — STAT5A	no	7925280, 10358079, 11413148, 16189514

Interaction	validated	PMID(s)
SRC — SYK	no	7513017, 12522270, 16189514
SRC — YES1	yes	11039464
STAT1 — FYN	yes	9804857
STAT1 — PTPN11	yes	12270932
STAT1 — SRC	yes	10358079
STAT1 — STAT3	yes	11594781
STAT1 — STAT5A	yes	10358045
STAT1 — SYK	no	10918587, 11294897
STAT3 — BTK	yes	17367410
STAT3 — CBL	no	15657067
STAT3 — CBLB	no	15657067
STAT3 — FER	yes	10878010
STAT3 — GRB2	no	15657067
STAT3 — HCK	yes	12244095
STAT3 — JAK3	no	8272872, 8626374, 9083098, 9305919, 9343414, 9566874, 9582023, 9872331, 10037026, 10446219, 10521505, 10918587, 11294897, 11335711, 11350938, 11429593, 11850821, 11940572, 12244095, 12576423, 12626508, 12763138, 14551213, 15465816, 15592455, 16189514
STAT3 — PTPN11	yes	11594781
STAT3 — PTPN6	no	15870198
STAT3 — SRC	yes	8657134, 12244095
STAT3 — STAT1	yes	11594781
STAT3 — STAT5A	yes	9398404
STAT3 — SYK	no	8272872, 8626374, 9083098, 9305919, 9343414, 9566874, 9582023, 9872331, 10446219, 10521505, 10918587, 11294897, 11335711, 11350938, 11429593, 11850821, 11940572, 12244095, 12576423, 12626508, 12763138, 14551213, 15465816, 15592455, 16189514
STAT3 — YES1	no	15657067

Interaction	validated	PMID(s)
STAT5A — BTK	yes	11413148
STAT5A — CBL	no	12193575
STAT5A — CRK	yes	11097834
STAT5A — CRKL	yes	9837784
STAT5A — JAK3	yes	9047382
STAT5A — PTPN11	yes	10617656
STAT5A — SRC	no	7925280, 10358079, 11413148, 16189514
STAT5A — STAT1	yes	10358045
STAT5A — STAT3	yes	9398404
STAT5A — SYK	no	7925280, 11413148, 16189514
SYK — BTK	no	8629002, 8630736, 11226282, 11598012, 12522270
SYK — CBL	yes	8621719, 9535867, 15556646
SYK — CBLB	yes	10022120
SYK — CRKL	yes	11313252
SYK — FYN	no	9535867, 12522270, 16713569
SYK — GRB2	yes	10747947, 11964172
SYK — LYN	yes	7831290
SYK — PIK3R2	yes	15536084
SYK — PLCG2	yes	10469124, 10981967
SYK — PTPN6	yes	10747947
SYK — SH3BP2	yes	9846481
SYK — SRC	no	7513017, 12522270, 16189514
SYK — STAT1	no	10918587, 11294897
SYK — STAT3	no	8272872, 8626374, 9083098, 9305919, 9343414, 9566874, 9582023, 9872331, 10446219, 10521505, 10918587, 11294897, 11335711, 11350938, 11429593, 11850821, 11940572, 12244095, 12576423, 12626508, 12763138, 14551213, 15465816, 15592455, 16189514
SYK — STAT5A	no	7925280, 11413148, 16189514

Interaction	validated	PMID(s)
YES1 — CBL	yes	9525940
YES1 — CBLB	no	15657067
YES1 — CSK	no	16094384
YES1 — DOK1	yes	11071635
YES1 — FYN	yes	12640114
YES1 — GRB2	no	15657067
YES1 — LYN	no	1544885
YES1 — PIK3R2	no	17620599
YES1 — SRC	yes	11039464
YES1 — STAT3	no	15657067

Appendix I

Selbstständigkeitserklärung

Hiermit erkläre ich, diese Dissertation selbstständig und nur unter Verwendung der angegebenen Hilfen und Hilfsmittel angefertigt zu haben. Sämtliche praktischen Hilfen, die über die im Laborbetrieb üblichen hinausgehen, sind spezifisch benannt. Alle wörtlichen und sinngemäßen Übernahmen aus anderen Werken sind als solche kenntlich gemacht.

.....	.....
Ort, Datum	Arndt Großmann